

us-09-812-350-17.p2n.rge

ri Feb 13 10:59:37 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - nucleic search, using frame_plus_p2n model

February 13, 2004, 01:35:16 ; Search time 7307 Seconds

on: (without alignments)
5100.406 Million cell updates/sec

US-09-812-350-17

fact score: 4550
quence: 1 MNPKFHKHKTETATAHEL.....VKMKRIEIEDDDNEEMIED 911

ring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

arched: 2889711 seqs, 20454813386 residues

al number of hits satisfying chosen parameters: 5777422

imum DB seq length: 0

imum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
MODEL=frame+ p2n.model -DEV=xl
=/cgn2_1/USPTO spool_p/US09812350/runat_10022004_101515_13351/app query.fasta_1.1095
DB=genEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOPCL=0 -LOOPEXT=0
NITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
LOCALALIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US09812350 @CGN 1 1 3983 @runat_10022004_101515_13351 -NCPU=6 -ICPU=3
NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase :

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.cm.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.ov.*
- 22: em.or.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*

RESULT 1

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4550	100.0	3105	8	UI3949 Arabidopsis
2	4545	99.9	2736	6	AX505540 Sequence
3	4545	99.9	2736	6	AX589851 Sequence
4	4388	96.4	6376	8	AF218796 Arabidops
5	4388	96.4	5821	8	AC020579 Arabidops
6	4003	88.0	3049	8	SOY8B100HS
7	3896	85.6	3052	8	AF083343
8	3883.5	85.4	3084	8	AF133840
9	3840.5	84.4	3148	8	AF133281
10	3825	84.1	3107	8	AF097363
11	3822.5	84.0	3058	8	AF174433
12	3790	83.3	2821	8	AF083344
13	3691	81.1	145149	2	AC130611
14	3691	81.1	162429	2	AC132451
15	3658	80.4	3109	8	OSA316025
16	3657	80.4	3942	8	AF077337
17	2768	60.8	197976	8	ATCHRIV39
18	2768	60.8	206606	8	ATFCA1
19	2502.5	55.0	2045	8	AF083327
20	2313	50.8	153750	2	AC129010
21	2299.5	50.5	300300	1	AP005373
22	2298	50.5	7726	3	LDHSP100
23	2294	50.4	10744	1	AE009461
24	2293	50.4	3179	1	BSU251205
25	2293	50.4	10473	1	AE014477
26	2288	50.3	5997	3	LMCLPB
27	2286	50.3	145709	1	D90914
28	2286	50.2	1570	8	AY062596
29	2282	50.2	298425	1	AP005049
30	2264	49.8	342800	1	AP003598
31	2258.5	49.6	10382	1	AE007611
32	2245.5	49.4	43980	6	AX413014
33	2245.5	49.4	258650	1	AL596171
34	2245.5	49.4	349980	6	AX417047
35	2243.5	49.3	303671	1	AE015937
36	2242.5	49.3	113316	2	AC007866
37	2242.5	49.3	159717	2	AC008368
38	2241	49.3	304708	1	AE017001
39	2238	49.2	301150	1	AP003190
40	2236.5	49.2	295050	1	AL591982
41	2236.5	49.2	349980	6	AX641671
42	2236	49.1	300950	1	AP005940
43	2234.5	49.1	38644	8	SPBCL6D10
44	2230	49.0	302050	1	AL935257
45	2228	49.0	290029	1	AE017027

ALIGNMENTS

ATU13949 3105 bp mRNA linear PLN 06-JUN-1995
Arabidopsis thaliana heat shock protein ATHSP101 (Athsp101) mRNA,
complete cds.
U13949
U13949.1 GI:537445
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 3105)
Schirmer, E.C., Lindquist, S. and Vierling, E.
An Arabidopsis heat shock protein complements a thermotolerance
defect in yeast
Plant Cell 6 (12), 1899-1909 (1994)
95170291
7866032
2 (bases 1 to 3105)
Vierling, E.
Direct Submission
Submitted (24-AUG-1994) Elizabeth Vierling, Biochemistry,
University of Arizona, Life Sciences South, Tucson, AZ 85721, USA
Location/Qualifiers
1. 3105
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="1-18"
/tissue_type="leaf"
/clone_lib="lambda ZAP (Stratagene) cDNA library from heat
shock poly(A) RNA; K.W. Helm and E. Vierling, Nuc. Acids
Res. 17:7995 (1989)"
/dev_stage="adult"
1. 3105
/gene="Athsp101"
<1.162
/gene="Athsp101"
163.2898
/gene="Athsp101"
/note="this protein is a member of the highly conserved
HSP100/Clp family of proteins, and like other HSP100/Clp
proteins, ATHSP101 has two consensus ATP-binding domains"
/codon_start=1
/product="ATHSP101"
/protein_id="AAA67927.1"
/db_xref="GI:537445"
/translations="MPEKFTKNTETIATAHELVNAGHAQFTPLHLAGALISDPTG
IFPQAISSAGGENAQAASERVINQALKLPSPQPPDDIPASSSLIKVTRAAQAKS
RGDTHLAVQLINGLEDQIOLLNEGVATARVSEVEKLRGEGKVESAGDTN
FOALTYGRDLVPAQGLDPVIGRDEIRRVVRLSRRTNNPVLIGFVGKTAIVE
GLAQRIVGVNPSLTDVRLISIDMGALVAGAKGEFEERLSVLKEVEDAEKVL
FIDEHLVLCAGTSGMDAANLKPMLARGQLRCI GATTEYRYKVEKDAFERF
QYVVAEPSVDPDTIILRGLEKVEYGHGVRIQDRLANAAQLSARYITGRHLPKAI
DLVDEACANRVQLDSPEIDNLKPKMOLTELHALEREKDKASKALIEVRKELD
DLRDLQPLTKYRKEKERIDEIRRLKQREELMFSIQEAREYRLARADRLRYGAIQ
EVESAIAQEGTSSEENMUTENGVPEHIAEVSVRWGTIPVTLGQNEKRLIGADR
LHKRVGQONQANVASRAILSRAGRAQOQTSGLFPGTGVGKTELAKALAEOLF
DDENLLVRIDMSEVMEQSVSLI GAPGVYGEHGGQTEAVRRPYCVILFDEVEK
AHVAVFTLLQVLDGDLRQDQGRVDFRNSVIMTSLNGLAEHLVLAGLTKGVMEVAR
DVMVREKRPPELLRLDEIVDFDPLSHDLRKVRLQMDVAVRLAERGVAVLAT
DADLYIILAESIDPVGARPIRRMWEKKVTELSKMVIREIDENSTVYIDAGAGLV
YRVESGLVDASTGKSDVLIHANGPKRSDAAQAVKMKRIEIEDDDNEEMIED"
2899.->3105
/gene="Athsp101"
3105
/gene="Athsp101"
/note="19 A nucleotides"
891 a 541 c 852 g 821 t

Alignment Scores: 7.12e-197 Length: 3105
Pred. No.: 4550.00 Matches: 911
Score: 100.00% Conservativity: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 8 Gaps: 0
DB: 8
US-09-812-350-17 (1-911) x ATU13949 (1-3105)
QY 1 MetAenProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
Db 163 ATGAATCCAGAGAAATTCACACACAGACAAACGACAAATTCCTACAGCTCATGAGCTA 222
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
Db 223 GCTGTGAATGACGAGCATGCTCAATTCATCTCTCTTTGCAATTTAGCTGGTGTGATCTCT 282
QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
Db 283 GATCCACCGGTATATTTCTCAAGCAATCTCTAGTGGCGGTGGCGAGACGACGCTCAA 342
QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
Db 343 TCTGCTGAAGAGTGATCAATCAAGCCTTGAAGAAGCTTCTTCAATCTCTCCACCT 402
QY 81 AspAspIleProAlaSerSerSerIleLysValIleArgArgAlaGlnAlaAlaGln 100
Db 403 GATGATATTCACGAGGAGTCTAGTCTTATTAAGGTCTATTCGCTGCTCAAGCTCTCAG 462
QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 120
Db 463 AAGTCACGAGGTGATCTCAATTTGGCTGTTCACCACTGATATGCGTCTTCTTGAAGAT 522
QY 121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
Db 523 TCTCAATCAGGAGTTTGTGAAGAGTGGGTAGCAGCGCGAGGGTAAAGCTCTGAG 582
QY 141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
Db 583 GTTCGAAGCTTCGTGGGAAAGAGGAGAAAGTGTAGAGTCTTTCGAGGAGAACGAG 642
QY 161 PheGlnAlaLeuLysThrThrTyrglyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
Db 643 TTTCAAGCTTTAAAGACTTATGAGAGAGATTTGGTTGAGCAGCAGGAGGAGTTGATCCT 702
QY 181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArgThrLys 200
Db 703 GTGATTGCTGTGATGAGGAGATTAGAAGAGTCTGTGAGGATTTCTTCGAGGAGAACGAG 762
QY 201 AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220
Db 763 AACATCTCTGTCTTATTTGGAGAGCCAGGAGTTGGTAAACAGCTGTGTGAGGAGTTTA 822
QY 221 AlaGlnAlaGlyValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
Db 823 GCACAAAGAGTTGTGAAGAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTTCG 882
QY 241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrglyGlyGluGluArgLeu 260
Db 883 TTGGACATGGGTGCTTGTAGTGTCTGTCTTAATCCAGGAGAGTTTGAAGAAGGTTG 942
QY 261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
Db 943 AAATCTGTTTTGAAGAAGTTGAGGACCTGAGGCAAGTATCTTCTTATTGATGAG 1002
QY 281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300
Db 1003 ATTCAATTTGTTTCTTGTGCTGCTGCAAACTGAAGGCTCGATGAGTGAAGTAACTGTTTC 1062
QY 301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrLeuGluGluTy 320
Db 1063 AAGCCCATGTTAGTAGAGGCGAGCTTTCATGATCTTGTGCTTACACGCTTGAAGATAC 1122

321	ArglySerValGluLysAspAlaAlaPheGluArgPheGlnGlnValTyrValAla	340	681	GluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspGlyArgLeu	700
1123	AGGAAATATGTGAGAAAGATGCTGCTTTGAGAGGAGGTTCACAGACTCAATGTTGG	1182	2203	GAGAAAGGCTCATGTTGCTGCTTCAACACTCTGCTCCAAAGTTTGGATGATGCTGATTG	2262
341	GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGly	360	701	ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleLeuMetThrSerAsn	720
1183	GAGCAAGTGTGCTGACACCAITAGTATCCTTAGAGGACTCAAGGAGAGATGATGAGGA	1242	2263	ACAGACGGGCAAGGAGGACAGTCTGATTTCAAGAACTCGGTGATANTCATGACATCAAC	2322
361	HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg	380	721	LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg	740
1243	CATCATGCTGTCGCAATCCAGACAGAGCTTTATATAATGCTGCTAGCTGTCTGCTGCT	1302	2323	CTTGTGCTGAACACTCTCTTCAGGGCTTAAGTGGAAAGTAACAATGGAAGTGGCCCGG	2382
381	TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla	400	741	AspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAsp	760
1303	TACATACTGTGTCGCAATTCAGGATTAAGCAATTAATTTGTTGATGAGGCTTGTGG	1362	2383	GACTGTGTGATCGGGAGGTGAGGAAACACTTCAGACAGAGCTCTTGAACAGGCTTCAAC	2442
401	AsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArg	420	761	GluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGln	780
1363	AATGTGAGAGTCCAGCTTGATGATCAACCTCAAGAGATTGATAACCTTGAAGCAAGG	1422	2443	GAGATTGTGCTGCTGACCCCTTTACATGACCATGACAGTTGAGGAAAGTAGCTCGGCTTCAA	2502
421	MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla	440	781	MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla	800
1423	ATGCAGCTGGAATTTGAATTCACGCTTGGAAAGGAGAGGATAAAGCCAGCAAGCT	1482	2503	ATGAAAGAGCTTGTGCTCGGCTTCTGTAAGAGAGAGTTGCTTTGGCAGTCACTGATGCT	2562
441	ArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThr	460	801	AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg	820
1483	CGACTTATAGAGGTGCGGAAGAGCTTGATGACCTGAGACAGCTTCAAGCTCTCACG	1542	2563	GCTTTGGACTATATCTTGGCAGAGATTATGACCCGCTGTTGTTGCTAGGCTTATAAGG	2622
461	MetLysTyrArgLysGluLysGluArgIleAspGluIleArgLeuLysGlnLysArg	480	821	ArgTyrMetGluLysValValThrGluLeuSerLysMetValValArgGluGluIle	840
1543	ATGAAATACAGAAAGGAGAGAGAGATTTGATGAGATTGCAAGGCTTAAACAGAAAGA	1602	2623	AGATGATGGAAGAGAGTGTGTGACAGACTGTCTCAAGATGTTGTTGCTGAGGAAATC	2682
481	GluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAlaArgAlaAla	500	841	AspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu	860
1603	GAAGAGCTCATGTTTCTTTCAGAGGAGCAGAAAGATATGACCTTGAAGAGCTGCT	1662	2683	GATGAAACTCCACTGTTTACATAGATGACGCGCTGCTGATCTTGTGTACCGGTAGAA	2742
501	AspLeuArgTyrGlyAlaIleGlnValGluSerAlaIleAlaGlnLeuGluGlyThr	520	861	SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla	880
1663	GATCTAGATATGCGCAATTCAGAGAGTGAATCTGCAATTTGCCAATTTGAAGGAAT	1722	2743	AGTGGAGGCTTAGTGGACGCTTCAACAGCAAGAGTCAGATGCTGCTCATATTGCT	2802
521	SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluVal	540	881	AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysMetArgIleGluGluIle	900
1723	TCTTCTGAAGAGATGTGATGCTCACAGAAACCTTGGGCTTGAACACATTTGCTGAGGTT	1782	2803	ACCGGCCCAAGAGAGTGTGATGCTCAGGCGGTGAGAGAGATGAGGATCGAGGAAATA	2862
541	ValSerArgTyrThrGlyIleProValThrArgLeuGlnAsnGluLysGluArgLeu	560	901	GluAspAspAspAsnGluGluMetIleGluAsp	911
1783	GTGAGCCCTTGGACAGGATTCAGTACGAGACTTGGCCCAAAATGAGAGGAGGTTG	1842	2863	GAGATGACGATAATGAGGAATGATCGAGGAT	2895
561	IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAla	580	RESULT 2		
1843	ATTGCTCTTGTGATAGCTTGCATAAGCGGCTTGTGGACAGAAATCAAGCGTAAATGCA	1902	AX505540	2736 bp	linear
581	ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly	600	LOCUS	Sequence 235 from Patent WO0216655.	PAT 27-SEP-2002
1903	CTTTCTGAGGCAATTTCTAAGTCAAGGCGAGGACTTGGTAGGCGCACACAGCCAACTGGA	1962	DEFINITION	AX505540	
601	SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla	620	ACCESSION	AX505540	
1963	TCATTTCTTATCTTGGACCAACTGCTGTTGGCAAACTGAGCTCGCCAGGCTTCTGCT	2022	VERSION	AX505540.1	GI:23386777
621	GluGlnLeuPheAspGluAsnLeuLeuValArgIleAspMetSerGluTyrMetGlu	640	KEYWORDS		
2023	GAGCAGCTGTTGATGATGAAACCTTCTAGTTCGATTGATGCTGGAATATATGAA	2082	SOURCE	Arabidopsis thaliana (thale cress)	
641	GlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGluGluGly	660	ORGANISM	Arabidopsis thaliana	
2083	CAACACTCTGCTCTCGCTCATTTGGGCGCACCCAGGCTATGTTGTTGCTCAGCAGAGG	2142	REFERENCE	Arabidopsis thaliana	
661	GlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGluVal	680	AUTHORS	Arabidopsis thaliana	
2143	GGACAACTAACTGAGGCTGTGAGGAGGCGACCTTATTGTGTCATACTCTTTGATGAAGTG	2202	TITLE	Arabidopsis thaliana	
			JOURNAL	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	
			REFERENCE	Harper, J.F.; Kreps, J., Wang, X. and Zhu, T. Stress-regulated genes of plants, transgenic plants containing same, and methods of use Patent: WO 0216655-A 235 28-FEB-2002; The Scripps Research Institute (US); Syngenta Participations AG (CH)	
			FEATURES	Location/Qualifiers	
			Source	1. .2736	
				/organism="Arabidopsis thaliana"	
				/mol_type="genomic DNA"	
				/db_xref="taxon:3702"	
			BASE COUNT	784 a 785 g 686 t	

GIN

gnment Scores:

d. No.:

1.05e-196

2736

re:
U. NO..

4545.00

910

cent since

99.89%

0

t Local

99.89%

11

rv Match

99.89%

10

LY Mall:

95.036

00

09-812-350-17 (1-911) x AX505540 (1-2736)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
1 ATGAATCCAGAGAAATTCACACACAGACAAACAGACAAATTGCTACAGCTCATGAGCTA 60
21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
61 GCTGTGAATGCAAGCAATGCTCAATTCATCTCTTGCATTAGCTGGTGTGTTGATCTCT 120
41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
121 GATCCCAACCGGTATATTCTCCTCAAGCAAATCTCTAGTGCCTGCGTGCAGAACCGAGCTCAA 180
61 SerAlaGluAtrqValIleAsnGlnAlaLeuLysLeuProSerGlnSerProProPro 80
181 TCTGCTGAAGAAGTGTATCAATCAAGCTTGAAGAAGCTTCCTTCCAAATCTCTCCACCT 240
81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGln 100
241 GATGATATCCAGCGAGTTCCTAGCTTATTAAAGTCAATTCGTGCTCAAGCTGCTCAG 120
101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 140
301 AAGTCACGAGGTGATACTCATTTGGCTGTGTGACAGTTGAATTATGGGTCTCTTTGAAGAT 160
121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
361 TCTCAAAATCAGGGATTGTTGAACGAAGTCGGTGTACGACGCGAGGGTAAAGTCTGAG 420
141 ValGluLysLeuArgGlyLysGluGlyLysValGlnSerAlaSerGlyAspThrAsn 160
421 GTTGAAGACTTCGTGGGAAGAAGGGAAGAAAGTTGAGAGTGCCTCAGGGGACACAAAT 180
161 PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
481 TTTCAGCTTTAAAGACTTATGAAGAGATTGGTTGAGCAACGACGGGAAGCTTGATCCT 540
181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArgThrLys 200
541 GTCAATTGGTTCGTATCAGGAGATTAGAAGAGTCTGTAGGAGATTCTTTCAGAGAACGAAG 600
201 AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValGluGluLysLeu 220
601 AACATTCCTGCTTATTGAGAGCCAGGAGTTGGTAAACAGCTGTGGTTGAAGGTTTA 660
221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
661 GCACAAGGATTCGTAAGACGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAAATTTTCG 720
241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGluArgLeu 260
721 TTGACATGGGTGGTTAGTTGCTGGTGCTAAATACCGAGAGAGATTGAAGAAGGTTG 780
261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
781 AATATCTGTTGAAGAAGTTGAGGACGCTGAAGCGCAAAAGTGATTCTCTTTATTGATGAG 840
281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300
841 ATTCATTGGTTCTTGGTCTGGCAAACTGAAGGGTGCATGGAGTGCAGCTTAATCTGTTC 900
301 LysProMetLeuAlaArgGlyGlnLeuArgCvsIleGlyAlaThrThrLeuGluGluLysr 320

901 AAGCCATGTTAGTACGAGGCGACCTTCGATGCTTCAACGCTGAAGATAC 960
 321 ArgLysTyrValGluLysAspAlaPheGluArgPheGlnValTyrValAla 340
 961 AGGAATATGTTGAGAAAGATGCTGCTTTGAGAGGAGGTTCCAAACAGCTATGTTGCG 1020
 341 GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGly 360
 1021 GAGCAAGTGTGCTGACCACTTAGTATCTTACAGGACTCAAGAGAGAGATGAGGGA 1080
 361 HisHISGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg 380
 1081 CATCATGTGTGGCAATCAAGACAGAGCTCTTATAATGCTGCTCAGCTGTCTGCTCGT 1140
 381 TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla 400
 1141 TACATACTGCTGCGCATTTACCGGATAAAGCAATGATTTGTTGATGAGGCTTGTGCG 1200
 401 AsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluAlaGlyArg 420
 1201 AATGTGAGAGTCCAGCTTGATGCTCACTCAAGAGATGATTAACCTTGAAGAGAGAGG 1260
 421 MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerIleAla 440
 1261 ATGCAGCTGGAAATTTGAATTCAGCCCTTGAAAGGAGAGAGGATAAAGCCAGCAAGCT 1320
 441 ArgLeuIleGluValArgLysGluLeuAspLysLeuArgAspLysLeuGlnProLeuThr 460
 1321 CGACTTATAGAGGTGCGAAGAGCTTGATGACCTGAGAGAGATGATTAACCTTGAAGAGAGG 1380
 461 MetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgGluLysGlnLysArg 480
 1381 ATGAATACAGAAAGAGAGAGAGAGAGATTCATGATTCGAGGCTTAAACAGAAAGA 1440
 481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAla 500
 1441 GAAGAGCTCATGTTTCTTTCAGGAGGAGAGAGAGATGATTAACCTTGAAGAGAGCTGCT 1500
 501 AspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr 520
 1501 GATCTAAGATATGGCGCAATTAAGAGTGGAACTCTGCAATTTGCCCACTTGAAGAGAACT 1560
 521 SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluVal 540
 1561 TCTTCTGAAGAGATGTGATGCTCACAGAAAGCTTGGCCCTGGAACATCTGCTGAGGTT 1620
 541 ValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeu 560
 1621 GTGAGCCGTTGACAGGGATTCAGTGACGAGACTTGGCCCAAAATGAGAGAGAGAGGTTG 1680
 561 IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAla 580
 1681 ATTGCTTCTGCTGATAGTTGATTAAGCGGGTTGTGGACAGATCAACGCGTAAATGCA 1740
 581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
 1741 GTTCTGAGGCAATTTAAGTCAAGGCGAGGACTTGAAGGCCCAACACGCAACTGGA 1800
 601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
 1801 TCATCTTATTCCTTGGAACCACTGCTGTTGCAAAACTGAGCTCGCAAGGCTCTGCT 1860
 621 GluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGluTyrMetGlu 640
 1861 GAGCAGCTGTTTGTATGATGAAACCTCTTAGTTGGATGATGATGCTGGAATATGGA 1920
 641 GlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGluGluGly 660
 1921 CAACACTGCTGCTCTGCTCATTTGGGGCACCACGAGGATGTTGTTGTCACGAGAGGT 1980
 661 GlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGluVal 680

1981 GGACAACAACTAAGTGGCTGTGAGGAGGCGACCTTATTGTGTCTATCATCTCTTTGATGAAGTG 2040
 681 GluLysAlaHisValAlaValPheAsnThrIleLeuGlnValLeuAspAspGlyArgLeu 700
 2041 GAGAAGGCTCATGTTGCTGCTCTTCAACACTCTGCTCCAGGTTTGGATGATGCTCGATTG 2100
 701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSerAsn 720
 2101 ACAGAGCGGCAAGGAGGAGGAGCTGCTGATTTTCAGAACTCGGTGATTAATCATGACATCAAC 2160
 721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740
 2161 CTTGGTGTGAAACACCTCTCTGAGGCTTAACTGGGAAAGTAACAATGGAAGTGGCCCGG 2220
 741 AspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAsp 760
 2221 GACTGTGTGATCGGAGGTGAGGAAACACTTTCAGACCGAGCTCTTGNACAGGCTTGAC 2280
 761 GluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGln 780
 2281 GAGATTGTGTGTGACCCCTTTTCACATGACCAAGTTGAGGAAAGTAGTCTCGGCTTCAA 2340
 781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800
 2341 ATGAAAGACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
 801 AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820
 2401 GCTTTGAGCTATATCTTGGCAGAGATTATGACCGGCTGATGCTGCTGCTGCTGCTGCTGCT 2460
 821 ArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluLeu 840
 2461 AGATGATGAGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520
 841 AspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu 860
 2521 GATGAAACTTCCACTCTTTACATAGATGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2580
 861 SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla 880
 2581 AGTGGAGGCTTAGTGACGCTTCAACAGGCAAGAGTCAAGATGCTGCTGCTGCTGCTGCTGCT 2640
 881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluLeu 900
 2641 AACGGGCCAAAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700
 901 GluAspAspAspAsnGluMetIleGluAsp 911
 2701 GAAGATGACGATTAAGGAAATGATCGAGAT 2733

RESULT 4
 AF218796
 LOCUS
 DEFINITION Arabidopsis thaliana heat shock protein 101 (HSP101) gene, complete cds.
 ACCESSION AF218796
 VERSION AF218796.1 GI:6715467
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Arabidopsis thaliana
 Rukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 6376)
 Hong S.W. and Vierling, E.
 Mutants of Arabidopsis thaliana defective in the acquisition of tolerance to high temperature stresses
 Proc. Natl. Acad. Sci. U.S.A. 97 (8), 4392-4397 (2000)
 20226114
 PUBMED 10760305
 2 (bases 1 to 6376)
 Hong, S.-W. and Vierling, E.
 Direct Submission

2512 GAGAGGATAAGCCAGCAAGCTCGACTTATAGAGT-ATGATACAGCTTCTTTCTTC 2570
 445 ----- 445
 2571 ATCAGTAGATAGCGCTTCAATATAAGACTATTGTGCTTGTCTTGGAAATCTTGATGTT 2630
 445 ----- 445
 2631 AGCATCTGTTTTTGGCATGAGTCTTTGATATAATTTCTCAAACTATGAATCTGTAGGTG 2690
 446 ArgLysGluLeuAspLeuArgAspLysLeuGlnProLeuThrMetLysTyrArgLys 465
 2691 CGGAAAGAGCTTGATGACCTGAGAGCAAGCTTCAGCCTCTCACGATGAATACAGAAAG 2750
 466 GluLysGluArgLysLeuAspGluLeuArgArgLeuLysGlnLysArgGluLeuMetPhe 485
 2751 GAGAAAGAGAGATTTGATGAGATTCGAGGCTTAAACAGAAAGAGAGAGCTCATGTTT 2810
 486 SerLeuGlnGluAlaGluArgArgTyrAspLeuAlaAlaAlaAspLeuArgTyrGly 505
 2811 TCTTTGAGGAGCGCAGAACGACGATATGACCTTGCAAGAGCTGCTGATCTAAGATATGC 2870
 506 AlaIleGlnValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAsn 525
 2871 GCAATTCAAGAACTGGAATCTGCAATTTGCCCACTTGAAGGAACTTCTTCTGAAGAGAA 2930
 526 ValMetLeuThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThr 545
 2931 GTGATGCTCACAGAAACGTTGGCCCTGCAACATTTGCTGAGGTGTGAGCGCTGGACA 2990
 546 GlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAsp 565
 2991 GGGATTCAGTGCAGAGACTTGGCCAAATGAGAGAGAGGTGTGATTTGGTCTTGCTGAT 3050
 566 ArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIle 585
 3051 AGGTTCGATTAAGCGGCTTGGGACAGACTCAAGCGGTAAATGCATTTCTGAGGCAATT 3110
 586 LeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeu 605
 3111 CTAAAGTCAAGGCGAGACTTGGAGCCACACAGCCAACTGGATCATCTTATTCCTT 3170
 606 GlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAsp 625
 3171 GGACCAACTGGTGTGGCAAACTGAGCTCGCAAGGCTCTTGCTGAGCAGCTGTTTGTAT 3230
 626 AspGluAsnLeuLeuValArgIleAspMetSerGluTyrMetGluGlnHisSerValSer 645
 3231 GATGAAACCTCTTAGTTGGATTGATATGCGGAATATATGGAACAACACTGTGTCTCT 3290
 646 ArgLeuIleGlyAlaProGly----- 653
 3291 CGCCTATTGGGCGCACCCAGGTAAGTAATCTATCTTTAGTTTTCATATCTTCTCA 3350
 654 -----TyrValGlyHis 657
 3351 TTATGAAATPATTTTTTCTTACTTACACGATTACTACACTTTCTCAGGTATGTGTCTAC 3410
 658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe 677
 3411 GAGAAAGGTGGACAACATACTAGGCTGTGAGGCGGACCTTATTGTGTCACTACTTCT 3470
 678 AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp 697
 3471 GATCAAGTGAGAGAGCTCATGTTGTGCTCTTCAACACTCTGCTCCAGTTTGGATGAT 3530
 698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleLeuMet 717
 3531 GGTGATTTGACAGCGGCAAGCAGACAGTGCATTTTCAGGAACCTCGGTGATATCATG 3590
 718 ThrSerAsnLeuGlyAlaGluHisLeuAlaGlyLeuThrGlyLysValThrMetGlu 737

Db 3591 ACATCAAAACCTTGGTGTCTGAACACCTCTTTCGAGGGCTAACTGGGAAAGTAACAATGGAA 3650
 Qy 738 ValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsn 757
 Db 3651 GTGCCCGGAGCTGTGTGATGCGGAGGTGAGGAAACACTTCAGACCAGAGCTCTTGAAC 3710
 Qy 758 ArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAla 777
 Db 3711 AGGCTTGACGAGATTGTGTGTCGACCCCTTTCACATGACCAGTTGAGGAAAGTAGCT 3770
 Qy 778 ArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaVal 797
 Db 3771 CGGCTTCAAAATGAAGACGTTGCTGTCGGCTTCTGAAAGAGAGTGTCTTTGGCAGTC 3830
 Qy 798 ThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProVal----- 813
 Db 3831 ACTGATGCTGTTTGGACTATATCTTGGCAGAGATTATGACCGGTT- AAGTCTCTTCAC 3889
 Qy 813 ----- 813
 Db 3890 ATGATGACAGAAATGCATATCATTTGTGATCTCTCAGGCTGGTTGTTCTCTGTGACTGACTC 3949
 Qy 814 -----TyrGlyAlaArgProIleArgArgTyr 822
 Db 3950 TTGTGGTGTCTTCTTGTGTTTGTGGTTAGGTGTATGCTAGTCTATAGGAGATGG 4009
 Qy 823 MetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIleAspGlu 842
 Db 4010 ATGGAGAAGAGGTGTGTGACAGAACTGTCAAAGATGTTGTGCTGAGGAAATCGATGA 4069
 Qy 843 AsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGluSerGly 862
 Db 4070 AACTCCACTGTTTACATAGATGACGAGCGCTGCTGATCTTGTGTACCGGTAGAAAGTGA 4129
 Qy 863 GlyLeuValAlaAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAlaAsnGly 882
 Db 4130 GGTCTAGTGGACGCTTCAACAGGCAAGATCAGATGCTGCTGATTCAATTTGCTAACCGG 4189
 Qy 883 ProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIleGluAsp 902
 Db 4190 CCAAGAGAACTGATGACGCTCAGCGGTGAAGAGATGAGGATCGAGGAAATAGAAGAT 4249
 Qy 903 AspAspAsnGluGluMetIleGluAsp 911
 Db 4250 GACGATATGAGGAAATGATCGAGGAT 4276

RESULT 5

AC020579 50821 bp DNA linear, PLN 19-JAN-2001
 LOCUS Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence,
 complete sequence.
 AC020579
 VERSION AC020579.5 GI:12324896
 KEYWORDS HTG
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 50821)
 AUTHORS Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,
 Barnstead, M.S., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
 Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence
 Unpublished
 REFERENCE 2 (bases 1 to 50821)
 AUTHORS Lin, X. and Kaul, S.
 JOURNAL Direct Submission
 REFERENCE 3 (bases 1 to 50821)
 AUTHORS Town, C.D. and Kaul, S.
 JOURNAL Direct Submission

JOURNAL
COMMENT

Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19 2001 this sequence version replaced gi:12280787.
Address all correspondence to:atc@tigr.org

BAC clone F1017 is from Arabidopsis thaliana chromosome 1

The orientation of the sequence is from SP6 to T7 end of the BAC
clone.

Genes were identified by a combination of several methods: Gene
prediction programs including GenScan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
of GlimmerX, see Mihaela Pertea,
http://www.tigr.org/softlab/glimmer.htm/glimmer.html, and
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpartea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/cgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).

Simple repeats are identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
Location/Qualifiers

TURES
source

```
1. 50821
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F1017"
/complement(9837..11757)
/gene="F1017.1"
/note="similar to GB:AA049376 from [Glycine max]"
complement(join(<9837..10159,10243..10343,10438..10485,
10576..10693,10800..10995,11127..11248,11404..11757))
/gene="F1017.1"
complement(join(10052..10159,10243..10343,10438..10485,
10576..10693,10800..10995,11127..11248,11404..11757))
/gene="F1017.1"
/codon_start=1
/product="putative choline kinase; 11757-10052"
/protein_id="AAG52400.1"
/db_xref="GI:12324898"
```

gene

mRNA

CDS

repeat_region

repeat_region

gene

mRNA

CDS

```
complement(12470..12501)
/rpt_family="AT rich"
complement(13045..13067)
/rpt_family="(A)n"
13093..16423
/gene="F1017.2"
/note="identical to heat shock protein 101 GB:AAF26423
from [Arabidopsis thaliana]"
join(<13093..14242,14327..14508,14649..15274,15359..15836,
15941..16423)
/gene="F1017.2"
join(13093..14242,14327..14508,14649..15274,15359..15836,
15941..16240)
/gene="F1017.2"
/codon_start=1
/product="heat shock protein 101; 13093-16240"
/protein_id="AAG52410.1"
```

```
/db_xref="GI:12324908"
/translation="WNPEKFTKWTETATATAHANWAGHAQFTPLHLAGALISDPTG
IPFALISSAGGENAQAASERVINQALKLPSPSPPPDDIPASSSLIKVIRRAQAQKS
RGDTHLAVDQLMGLEDQIDRLNEVGAVATARKSEVEKURKGGKGVSAAGDTN
FOALKTYGRDLVQOAGKDPVIGRDEIRRVVRIILSRRTKNNPVLIGEPGVGATVE
GLAQRITGVDPNSITDRLIISLDMGALVAGAKYRGEFERKSLVKEGDAEGKVL
FIDEHLVLAGAGTEGSDAANLFPMLARGQLRCIGATTLEEKYRKYVKEDAFFERRF
QQVVAEPSVPTTISILRGLKEKYGHHGVRIQDRALINAAOLSAARYITGRHLPOKAI
DLVDEACANRVQLOSDPREIDNLERKMQLEIELHALEREKDKASKARLIEVKELD
DURDKLOPTMKYRKEKRIEIRLQKREBELFSLQAEIRRYDILARADLRIGAIQ
EVESALLEGTSSENNMLTENVEGHTARVSWNTGIPVTRLGQNEKERLIGADR
LHKRVVQNAVNSAEALRSRAGIGRPGQPTGSGFLGTCVGTGKTELAKALBOLF
DENLIVRDMSEYVBOHSRLICAPPGYVGHGGGQLTEAVRRPYCVILFDEVEK
AHVAVNTLLOVLDGRLTDGQRTVDGQRTVDNSVIINTSNLGAHLLAGLTGKVTMEVAR
DCVMREVRKHFEPELLRLDEIVDPDLSHDQLRKVARLOKMDVAVRLAERGVAVAT
DAALDILAESDPVYGARPIRRWMEKKVITELSKMVVEEDENSTVTLAIDGADLV
YRVESGLVDASTGKSDVLIHANGPKRSDAAQAKKRIEIBDDDDNEEMIED"
complement(16481..17587)
/gene="F1017.3"
/note="similar to unknown protein GB:AAD21437 from
[Arabidopsis thaliana]"
complement(join(<16481..17251,17318..17587))
/gene="F1017.3"
complement(join(16481..17251,17318..17587))
/gene="F1017.3"
/codon_start=1
/product="unknown protein; 17587-16481"
/protein_id="AAG52411.1"
/db_xref="GI:12324909"
/translation="NAIGPSSVIIIVGIIILATYOSILKPPPPNLCGSPGPIITAPR
IKLRDGRHLAYEYGLPRKAKHIVFIHGSDCRHDVAFATLLSPDLVQERGVTVMS
FDKPGYGBSDPPIRTPKSLALDIBELADQSLGSKFYVIGKMGQAQAGCLKVTPEH
RLAGVTLVAVVNYWRNLPLNISTEGENLQQRDQWAVRVAVAHVPMLYWNTQWFE
PGSSVNRDGGVLSQPDKDIILKLGSSRXPFLAEVROQGIHESINRDMIVGFGNWEFD
PLELENPFLNREGSVHLWQGDMLVPVTLQRYIADKLPLWHLHYHVGSGSHFPLAPG
VVDEIVKTLNHH"
complement(18687..21119)
/gene="F1017.4"
/note="similar to unknown protein GB:AAD21437 from
[Arabidopsis thaliana]"
complement(join(<18687..19284,20351..20547,20799..21119))
/gene="F1017.4"
complement(join(18687..19284,20351..20547,20799..21119))
/gene="F1017.4"
/codon_start=1
/product="unknown protein; 21119-18687"
/protein_id="AAG52412.1"
/db_xref="GI:12324910"
/translation="MASDSSRXSLKPSRSKSLFPFVVIVIGIIVALTYSOKLKP
QPKLCCSSGPIITAPRIKLDGRVLAKEHGLPREKANRKIVFIHGSDCRHDVAF
TLLSPDLVEELGVMSYPRPGYCSDDPHSPTRSLVSDIBELDQSLGSKFYVIG
KMGQAQAGCLNLYIPEHLAGVTLVAVVNYWRNLPLNVSTEGFNQQRDQWAV
RVAVAPMLIYWNVTQKXFPFSSIANRSLLSQSDRIISKRGYTRKPHWAEVRQGI
HESINRDMIVGFGNWEFDIDIDNEFLANNEGFLWQGDMLVPVKQRYLAHQLPW
VHYHEVRSRGHFFHTKGVNDIVITVLTITDTDTIRQ"
complement(21829..23197)
/gene="F1017.5"
/note="similar to unknown protein GB:AAD21437 from
[Arabidopsis thaliana]"
complement(join(<21829..22626,22877..23197))
/gene="F1017.5"
complement(join(21829..22626,22877..23197))
/gene="F1017.5"
/codon_start=1
/product="unknown protein; 23197-21829"
/protein_id="AAG52399.1"
/db_xref="GI:12324897"
/translation="WTRDSSRNSSLKPMWRKFLPSPVIVIGIIVALTYSOKLKP
PKLCCSSGPIITAPRIKLDGRVLAKEHGLPREKANRKIVFIHGSDCRHDVAF
TLLSPDLVEELGVMSYPRPGYCSDDPHSPTRSLVSDIBELDQSLGSKFYVIG
YMGQAQAGCLKLYIPEHLAGVTLVAVVNYWRNLPLNVSTEGFNQQRDQWAV
AHYTPMLIYWNVTQKXFPFSSIANRDSLLAQDKDIISKRGYTRKPHWAEVRQGIH
ESINRDMIVGFGNWEFDIDNEFLANNEGFLWQGDMLVPVKQRYLAHQLPW"
```

gene	HYHEVPRSGHFFHYTKGVVDDIVKSLTSDVRAASRVC1	Db	13573	TTTCAAGCTTTAAAGACTTATGGAAGAGATTGGTTGAGCAAGCAGGAAAGCTGTATCT	13632
	/gene="F1017.6"				
	/note="similar to ribosomal protein L33B GB:NP_014877 from [Saccharomyces cerevisiae]"	Qy	181	ValIleGlyArgAspGluGluLeuArgValValArgIleLeuSerArgArgThrLys	200
mrna	join(<23489..23520,23750..23785,24176..24404,24499..24554)	Db	13633	GTGATTGGTGGTATGAGGAGATTAGAAAGAGTGGTGGAGGATCTTTCGAGGAGAACGAAG	13692
	/gene="F1017.6"				
CDS	join(23489..23520,23750..23785,24176..24404,24499..24540)	Qy	201	AsnAsnProValLeuLeuGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu	220
	/codon_start=1	Db	13693	AACAATCTGTGCTTATTGGAGAGCCAGGAGTTGGTAAACACAGCTGTGGTTGAAGGTTTA	13752
	/product="putative ribosomal protein; 23489-24540"				
	/protein_id="AAG52401.1"	Qy	221	AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuLeuSer	240
	/db_xref="GI:12324899"	Db	13753	GCACAAAGAGATTGAAAGAGAGATGTCACCAACAGCTTACTGATGTGAGATTAATTCG	13812
	/translation="WVKGQGRVRLVYRGTTILGYKRKSNQYPTSLVQIEGVNTQF EVNWKGRMAVIYRAKTKNGSHYRCIWKVTRPHGNSGVVRAKFTSNLPKSGSR VRFVWYPSNI"	Qy	241	LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGluArgLeu	260
gene	25509..29950	Db	13813	TTGACATGGGTGGTGTAGTTGCTGGTCTAAATACCGAGAGAGTTTGAAGAAAGTTG	13872
	/gene="F1017.7"				
mrna	join(<25509..25911,25995..26563,26663..29047,29144..29950)	Qy	261	LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu	280
	/note="similar to GB:K1AA0361 from [Homo sapiens]"	Db	13873	AAATCTGTTTGAAGAAGTTGAGGACGCTGAAGCAAGCTGATTCTCTTTATTGATGAG	13932
	/gene="F1017.7"				
CDS	join(25509..25911,25995..26563,26663..29047,29144..29950)	Qy	281	IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe	300
	/codon_start=1	Db	13933	ATTCAATTGGTTCTTGGTGTGTCGCAAACTGAAGGGTCGATGGATGCAGCTAATCTGTT	13992
	/product="putative phosphoribosylformylglycinamide synthase; 25509-29950"	Qy	301	LysProMetLeuAlaArgGlyGluLeuArgCysIleGlyAlaThrThrLeuGluGluTyr	320
	/protein_id="AAG52403.1"	Db	13993	AAGCCCATGTTAGCTAGAGGCGAGCTTCGATGATGATGGTCTACACGCTTGAAGATAC	14052
	/db_xref="GI:12324901"				
	alignment Scores:				
	Seq. Id. No.: 2.78e-188 Length: 50821	Qy	321	ArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyrValAla	340
	Score: 4388.00 Matches: 908	Db	14053	AGGAAAATATGTTGAGAAAGATGCTGCTTTGAGAGAGGTTCCAAAGTCTATGTTGCG	14112
	Percent Similarity: 86.56% Conservative: 0	Qy	341	GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGly	360
	at Local Similarity: 86.56% Mismatches: 3	Db	14113	GAGCAAGTGTGCTGACACCATTAGTATCTTCTTAGAGGACTCAAGGAGAGATAGGGA	14172
	Every Match: 96.44% Indels: 140	Qy	361	HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg	380
		Db	14173	CATCATGGTGTGGATCCAAAGACAGAGCTCTTATAATGCTGCTCAGCTGCTGCTCGT	14232
		Qy	381	TyrIleThr	383
		Db	14233	TACATAACTGGTATGTTAAGATTCTTAATCCTAAGCTGATGTTTATGTTTCATATAGTG	14292
		Qy	384	-----GlyArgHisLeuProAspLysAlaIle	392
		Db	14293	TTGTTCTTGACATGACATTTTGTGGGATTTAGTGGCGATTTCCGGATAAAGCAATT	14352
		Qy	393	AspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGlu	412
		Db	14353	GATTTGGTTGATGAGGCTTGTGCGAATGTCAGAGTCAGCTTGTATGATCAACCTGAAGAG	14412
		Qy	413	IleAspAsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArg	432
		Db	14413	ATTGATAACCTTGAAGGAAAGAGAGTGCAGCTCGAAATTTGAACCTTCACGCTTGGAAAG	14472
		Qy	433	GluLysAspLysAlaSerLysAlaArgLeuIleGluVal	445
		Db	14473	GAGAAGGATAAGCCAAAGCTCGACTTATAGAGGT-ATGATACAGCTTTCTTTCTC	14531
		Qy	445	-----	445
		Db	14532	ATCAGTAGATAGCGCTTCAAAATATAAGACTATTGCTGCTTCTCGAAATTTCTGATGTT	14591
		Qy	445	-----	445
		Db	14592	AGCATCTGTTTTCGCGATGAGTCTTTGATATAAATTTCTCAAACTATGATCTGTAGGTG	14651
		Qy	446	ArgLysGluLeuAspLeuArgAspLysLeuGlnProLeuThrMetLysTyrArgLys	465

Qy	798	ThrAspAlaAlaIeuAepTyrIleIeuAlaGluSerTyArgProVal-	----- 813
Db	15792	ACTGATGCTGCTTTGGCATATATCTTGGCAGAGAGTTATGACCCGGT-AAAGTCTCTTCCAC	15850
Qy	813	-----	813
Db	15851	ATGATCGCAAGTGCATATCATTTGTTGATCTCTCAGCGTGGTTTGTCTTGTGACGTGACTC	15910
Qy	814	-----TyrGlyAlaArgProIleArgArgTrrp	822
Db	15911	TTTGCTGCTTGCTTTCTTTGTTTCTTGTTAGGTGTATGTGCTAGGCCATTAAAGGAGATGG	15970
Qy	823	MetGluLysIysValValThrGluLeuSerIysMetValValArgGluGluIleAspGlu	842
Db	15971	ATGGAGAAGAAGTGGTGACAGAACTGTCAAAGATGGTTGTGCGTGAGAGAAATCGATGAA	16030
Qy	843	AsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGluSerGly	862
Db	16031	AACTCCACTGTTTACATAGATCGACGGCGCTGGTGATCTGTGTACCGGGTAGAAGTGGGA	16090
Qy	863	GlyLeuValAspAlaSerThrGlyLysIysSerAspValLeuIleHisIleAlaAsnGly	882
Db	16091	GGTCTAGTGGACGCTTCAACAGCCAAAGATCAGATGTGCTGATTTCATATTGCTAACCGG	16150
Qy	883	ProLysArgSerAspAlaGlnAlaValIysIysMetArgIleGluGluIleGluAsp	902
Db	16151	CCAAAGAGAAGTGATGACGCTCAGCGCGGTGAAGAAGATGAGGATCGAGGAATAGAAGAT	16210
Qy	903	AspAspAsnGluGluMetIleGluAsp	911
Db	16211	GACGATTAATGAGGAATGATCGAGAT	16237
RESULT 6			
SOYSB100HS			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
COMMENT			
FEATURES			
source			
gene			
CDS			

LFIDEHLVLGAGRTGEGMDANLFPKMLARGQLRCIGATTLEBYRYKVEKDAAFERR
FOOVFAEPSVDVDTIILRGLKERVEGHGVIOIDRALMAAOLSNRYITGRHLPDKA
IDLVDVACANVRVOLDSPBEIDNLERKEMQLEVEHLHLEKEDKAKARLVEYRKEL
DLDRKQOLVMKMYKREKREVEDEIRLKKKREELLFALQEABRYIDARAADLYKAI
QEVETAIQOLSGSTEENLMTETVPEQIAEVVSRWGTIPVTRIGQNEKERLIGLGR
LHRSVVGQDQAVNAVEAVALRSRAGLRPOQPTGSLFPGTGVKTELEAKALAEOLF
DNEOLVRIDMSYMECHSVSRLLGAPPGYVHBEQGLTEAVRRPYSVVLFPDEVRK
AHTSVENLLQVLDGRITDGGRTVDFRNTVITWISNLGAELHLLSGLSGKCTWQVAK
DRWQEVPRORPELNLDELIVDPDLSDHQLKVAELBLOKMDKVASRLAERKIALAVT
DHALDIILSESVDYFYGARPIRMLEKKVUTSELSEMLVREIDENSTYIIDAGNNGE
LVYRKEKNGVGNVPTGQKSDILIQIPNGHAPKTDVAQVKKMKLEIDDDDEMESE

BE COUNT 850 a 564 c 863 g 772 t

GIN

gment Scores:

d. No.: 3,32e-172 Length: 3049
Matches: 794
Conservative: 72
Similarity: 94.85%
at Local Similarity: 86.97%
Matches: 87.98%
Indels: 8
Gaps: 6

09-812-350-17 (1-911) x SOYSH100HS (1-3049)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
136 ATGAATCCTGAGAAGTTTACTCACAAGACTAATGAAGCTTGTGCGCGGCCACGAGCTC 195
21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLeuSer 40
196 GCGATGAGTTCAGGCGCACCGCAATTGATCCCATCCACTTGGCCCGCGCGCTGATTTCC 255
41 AspProThrGlyIlePheProGlnAlaIleSerSerAla--GlyGlyGluAsnAlaAla 59
256 GATCCCAACGGCATCTTCGTGCTAGCGATAACACGCGCGCGCGCGGCGGAGGATCGCA 315
60 GlnSerAlaGluArgValIleLeuGlnAlaLeuLysLeuProSerGlnSerProPro 79
316 CGCGCGTGGAGCGAGTGTGAACAGGCTCTGAAGAGCTACCTGCGCGCGCGCGCGCG 375
80 ProAspAspIleProAlaSerSerSerIleIleValIleArgAlaGlnAlaAla 99
376 CGGAGCGAGTGGCGGAGCAGCACCACTCTGTGAGGCGCATCAGGAGGACAGCGCGCG 435
100 GlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGlu 119
436 CAAATATCAGTGGCGACACGCGTGTGGCGGTGATCAGTTCATCTCGAATCTCGAA 495
120 AspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
496 GACTCCAAATCGAGACCTTGTGAAGAGCGGGGTTGCGGTGGCGGAGGATAGATCG 555
140 GluValGluLysLeuArgGlyLysGluLysGluLysValGluSerAlaSerGlyAspThr 159
556 GAAGTGATAGCTTCGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 615
160 AsnPheGlnAlaLeuLysThrTyGlyArgAspLeuValGluGlnAlaGlyLysLeuAsp 179
616 AATTTCCAGCTTTGAAGACTTATGCGCGTGACCTTGTGTAACAGCGCGGAGGAGTCCAC 675
180 ProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgThr 199
676 CCGTGTATTGGCGGTGACGAGAGATAGTAAGAGGTTGTGAGGATTCATCAGCGAGGACT 735
200 LysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGly 219
736 AAGACAAACCCGGTTCCTGTTGAGAGAACCGGGTGTGGGAAACCTCGGTTGTGGAGGG 795
220 LeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIle 239
796 TTGGCAGACAGGATAGTAAAGAGGGATGTTCCAAAGCAACCTTCTGATGTGAGGCTATT 855
240 SerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyArgGlyGluPheGluGluArg 259

856 GCCTTGGATATGGGGCGTGTGTCGGGGTCCCAAGTATAGAGCGGAGTTTGAGGAGCGG 915
260 LeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAsp 279
916 TTAAGGCTCTTTTGAAGGAAGTGGAGGAGGCTTGAGGGAGGAGGTATATCTTTCATTGAT 975
280 GluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeu 299
976 GAGATTCTATTGCTCTTGGTGTGTGAGAACTTGAAGGCTCCATGATGCTGCTATCTA 1035
300 PheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrLeuGluGlu 319
1036 TTCAAACTTATGCTTGTGCGCGCCAGCTTAGTGTGCTTGGTCCACCGCTTGAGGAG 1095
320 TyrArgLysTyValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyVal 339
1096 TACAGGAAGTATGAGGAAGGATGCTGCATTCGAGGAGGAGTTCACACAGGTTTTTGIG 1155
340 AlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlyLysTyGlu 359
1156 GCGGAACCTAGTGTGTTGATACCATTCATTCCTGCTGCTTGAAGAGAGATATGAA 1215
360 GlyHisGlyValArgIleGlnAspArgAlaLeuLeuLeuAlaGlnLeuSerAla 379
1216 GCGCATCACGCTGTAGAAATTCAGGACCGCTGCTTGGTATATGCGAGCTCAATTTGTAAAC 1275
380 ArgTyIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCys 399
1276 CGGTATATACTGGCGCTCATCTCTGACAGGCAATTCATTTGTTGACAGGCTTGT 1335
400 AlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLys 419
1336 GCAATGTAGGCTTCACTTGTATGATGCTGAGGAAATTCATAACCTTGAAGAGAG 1395
420 ArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysValSerLys 439
1396 AGAATGAGCTAGAGTGAAGTTCATGCTCTGAGAAAGAGAAAGACAGGCTAGCAAA 1455
440 AlaArgLeuIleGluValArgLysGluLeuAspAlaLeuArgAspLysLeuGlnProLeu 459
1456 GCGCGCTTGTGAGTGGAGGAACTTGTATGATCTTGAGGAGCAAGCTTTCAGCCTTIG 1515
460 ThrMetLysTyArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLys 479
1516 ATGATGAAGTACCGAAAGAGAGAGAGGCTGATGAGATTCGAAGGCTCAAGAGAAA 1575
480 ArgGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyArgAspLeuAlaArgAla 499
1576 AGAAGAGCTTCTTTTGTCTTACAGAGGCTGAGAGAGATATGATCTGGCTAGAGCT 1635
500 AlaAspLeuArgTyGlyValIleGlnGluValGluSerAlaIleAlaGlnLeuGluGly 519
1636 GCACACCTGGATATGAGCAATTCAGAGGTGGAACCTGCAATACCAACTTGAAGGG 1695
520 ThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu 539
1696 ---ASGACTGAAGAGAAATCGATGTTGACTGAAACTGTTGGACCGGAGCAATAGCTGAG 1752
540 ValValSerArgTyThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArg 559
1753 GTTGTGAGCGCTGGACCGGTATACCAAGTTCAGAGGCTTGGCCAAATGAAAGAGAGA 1812
560 LeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsn 579
1813 TTGATTGAGCTTGTGAGAGATTCACAGCAGCAGAGTTGTAGGACAGCAAGCTGTATAT 1872
580 AlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThr 599
1873 GCTGTGCTGAGGCTGTTGATGATCTAGAGCTGGCTAGAGAGCTCAACCAACCAACT 1932
600 GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeu 619


```

1933 GGTTCATTCCTATCTCTGGTCCAACTGGTGTGGTAAACCTGAGCTTGCAAGGCTCTT 1992
620 AlaGluGlnLeuPheAspGluLeuValArgIleAspMetSerGluTyrMet 639
1993 GGTGAGCAACTATTTGATGATAAACCAGTTGGTGAAGATATGATGTCGAAATACATG 2052
640 GluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGluGlu 659
2053 GAGCAGCATCGTTTCAGATTGATGTTGACCAACAGGATGTTGGACATGAGGA 2112
660 GlyGlyGlnLeuThrGluAlaValArgArgAProTyrCysValIleLeuPheAspGlu 679
2113 GGTGGCAACTAACTCAAGCGGTGAGCGGATGAGCACTTACAGTGTGTGCTTTTGTATGAA 2172
680 ValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspGlyArg 699
2173 GTGGAAGGCAATCATCTGTGTTCAATATCTCTTCAAGTTTGGACCATGGAGG 2232
700 LeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSer 719
2233 TTAAGTATGAGCAAGCGGTACTGTGGACTTTAGAACACTGTAATATCATGACCTCA 2292
720 AsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAla 739
2293 AATCTTGGAGCAGACCACTCTCTCAGTGGACTTTACAGGAAATGTACCATGCAAGTTGCT 2352
740 ArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeu 759
2353 CGGGATCGGGTAATGAGGAGGTGAGAGGCAATTTAGGCCAAGATTTGCTGATGACTT 2412
760 AspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeu 779
2413 GATGAATTTGTGATTTGACCCCTTTACATGACCAATTTGAGGAAGTTGCAAGGCTA 2472
780 GlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAsp 799
2473 CAAATGAAGCATGTAGTAGTCTCTTCTGTCGAGAAAGAAATTTGCTTGGCAGTCACTGAT 2532
800 AlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIle 819
2533 GCAGCACTAGACTATATATCTTCCGAGAGCTATGATCCGGTGTATGGTGTAGACCAATA 2592
820 ArgArgTyrMetGluLysValValThrGluLeuSerLysMetValValArgGluGlu 839
2593 AGAGGTGGCTTGAAGAGAGGTGGTGCACAGATTTGTCTAGATGCTTGTAGAGAGGAG 2652
840 IleAspGluAsnSerThrValTyrIleAspAlaGly-----AlaGlyAspLeuValTyr 857
2653 ATTGATGAGAATTCACCGTTTACATAGATGCTGGACCCCAATGGGGCGGATTTGCTTAC 2712
858 ArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysSerAspValLeu 876
2713 CGTGTAAGAAAGATGAGGGGTGTTTAACTTACACTGGCAGAGATCCGATATCTTGT 2772
877 IleHisIleAlaAsnGly-----ProLysArgSerAspAlaAlaGlnAlaValLysLys 894
2773 ATTCAGATACCTAATGGACATGCACCTAAA---ACTGATGCTGTTCAAGCAGTCAAGAAG 2829
895 MetArgIleGluGluIleGluAspAspAsnGluGlu 907
2830 ATGAAGATTGAGGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2868

```

SULT 7
083343
CUS
FINITION
CESSION
RSION
YWORDS
NURCE
ORGANISM

```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 3052)
Wells,D.R., Tanguay,R.L., Le.H. and Gallie,D.R.
HSP101 functions as a specific translational regulatory protein
whose activity is regulated by nutrient status
Genes Dev. 12 (20), 3236-3251 (1998)
99003076
9784498
2 (bases 1 to 3052)
Gallie,D.R., Tanguay,R.L. and Wells,D.R.
Direct Submission
Submitted (10-AUG-1998) Biochemistry, University of
California-Riverside, Riverside, CA 92521-0129, USA
3 (bases 1 to 3052)
Gallie,D.R., Tanguay,R.L. and Wells,D.R.
Direct Submission
Submitted (06-DEC-2000) Biochemistry, University of
California-Riverside, Riverside, CA 92521-0129, USA
Sequence update by submitter
On Dec 6, 2000 this sequence version replaced gi:3902830.
Location/Qualifiers
1..3052
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/db_xref="taxon:4097"
1..3052
/gene="HSP101"
101..2830
/gene="HSP101"
/codon_start=1
/product="101 kDa heat shock protein"
/protein_id="AAC83688.2"
/db_xref="GI:11561806"
/translation="MNPEKETHKNEALACALALASAGHAQFTPLHMAVALISDRHG
IFRQAVNAGNBEVANSVERLNQNMKLPSPQTPADEIPPTSLIKVLRAQSSOK
SRGDSHVALQLILGLLEDSQIGDLKEAGVSASRVKSEVEKLRGKRGKVESASGDT
TFQALNTYGRDLNVEQAGKLDVIGRDEIRRVRIILSRRTKNNPVLIGEPGVGTAVV
EGLAORTVRGDPNLDVRLIALDMGALVAGARGEPERKAVLKEVEEAEKVI
LFIDEHLVLCAGRTGEGMDAANLFPKMLARGQLRCIGATTLEBYRYVVKDAFRR
FQVYVAEPVADTISILRGKERYEGHGVKIQDRALVVAALSSRYITGRHLPDKA
IDLVDCAANRVQDSPEIDNLERKIQLEVEHLHALEKDKKASKALIEVRKEL
DLDRLQLPMYKKEKIDELRLKRDDELIALQAEARRYDLARAADLURIGHI
QEVETALNTLESTSAETMLTETVPDQIAEVSRTWTGIPVSRIGQNEKELIGLGR
LHQRVGDHVAARAVALRSRAGLGRPOPTGSFLPGTGVGTETELAKALAEQLF
DDDKLMLRDSMEVMEQHSVSRILGAPGYVGHDEGGQLTEAVRRRYSVVLDEVEK
AHPAVNTLQVLDGRLTGCGRTVFTNSVILMTSNLGAELVLSGLMGKCTMETAR
EMWQVRKQPKPELLRLDEIVVDFLHKLQKQVCRVQMKOVALRALRGALGVT
EAPDVLITSDYPVYGARFIRLWLRKVTLSKMLVKEIDENSTVIIDAGVSGKD
LTYRVEKNGGLVNAATGQKSDILQLPNGPRSDAVQAVKVRIBEEIDEMED"
BASE COUNT 898 a 536 c 836 g 782 t
ORIGIN

Alignment Scores:
Pred. No.: 2.23e-167 Length: 3052
Score: 3896.00 Matches: 774
Percent Similarity: 93.19% Conservative: 75
Best Local Similarity: 84.96% Mismatches: 56
Query Match: 85.63% Indels: 6
DB: 8 Gaps: 5

US-09-812-350-17 (1-911) x AF083343 (1-3052)

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
Db 101 ATGAATCTGTAATAATTCCACCACAGACTAACAGGCCCTTGTGGGCACTCGAGCTA 160
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
Db 161 GCATATCCGCGAGGCGATGCTCAATTTACGCTCTGCATATGCTGTGGCTTAAATATCT 220
QY 41 AspProThrGlyLysPheProGlnAlaIleSerSerAlaGlyGly---GluAsnAlaAla 59

```

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
FEATURES

source
gene
CDS

221	GATCA	CAATGGTATTTTTCGACAGCGGATTTGCTCAATCCTGGTGGGAATCGAGAGTACT	280
60	GlnSer	AlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro	79
281	AA	TTCAAGTGGAGCGGTATTTGAATCAAGCGATGAAGAAGCTACTCTTCTCAAAACCGGCT	340
80	ProAsp	AspIleProIleAsnSerSerLeuIleLysValIleArgArgAlaGlnAla	99
341	CCTG	CGAAATCCCACTAGCACTTCCCTTATCAAGGTGTATCGCGAGCAAACTCGTCG	400
100	GlnLys	ArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGlu	119
401	CAGA	AGTCTCGTGGTGACACCAATTTAGCAGTGGATCAGTTGATTTTAGSACTGCTAGAA	460
120	AspSer	GlnIleArgAspLeuAsnGluValGlyValAlaThrAlaArgValLysSer	139
461	GACT	CCCAATTTGGAGATCTTTTGAAGAAGCTGGAGTGATGCTCAAGAGTGAATCA	520
140	GluVal	GluLysLeuArgGlyValGlyLysLysValGluSerAlaSerGlyAspThr	159
521	GAGT	AGAGAACTTAGAGAAAGGAAGAAGAAGTCGAAAGTGTCTCAGGGACACC	580
160	AsnPhe	GlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAsp	179
581	ACAT	TCCAAGCACTCAACACTTATGCCGTGATCTTGTGGAAACAAGCAGGAAGCTTGAT	640
180	ProVal	IleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArgThr	199
641	CCCG	TGATGGTAGGATCAAGAATTAAGAAGTCGTTCGGATTTTATCAAGAGGACT	700
200	LysAsn	AsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValAlaGluGly	219
701	AAGB	ACACCCGGTCTTATTGGAGAGCCGGTGTGGTAAACAGCAGTTGTTGAAGGG	760
220	LeuAla	GlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIle	239
761	CTAG	CAAGAGGATGTAGCTGTGTATGTCAGATTAATTTAGCTGAATTTAGCTTATA	820
240	SerLeu	AspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGluArg	259
821	GCAT	TGATATGGAGCGCTAGTTCTGGAGCTAGTACAGAGTGAATTTGAAGAGAGG	880
260	LeuLys	SerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAsp	279
881	CTGA	AGCGTGTCTGAAAGAAGTTGAAGACGGGAAGGAAGTAATCTTTTCATTGAC	940
280	GluIle	HisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeu	299
941	GAGAT	ACATTTAGTCTCGGTGCTGTGGACAGAGGCTATAGATGTCTGCTAATCTG	1000
300	Phelys	ProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGlu	319
1001	TTTA	AGCCATGTCTAGCCAGGTCATTTACCGTGCANTTGGTGGCAATACACTCGAGGAG	1060
320	TyrArg	LysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnValTyrVal	339
1061	TACAG	AAGTATGTTGAGAGGATGCTGCATTTGAGAGCGCTTTCCAGCGGTGTATGTT	1120
340	AlaGlu	ProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGlu	359
1121	GCTG	AGCCTAGTGTGTGACACTATTAGTATTCTCCGTGGGTTCGAAGGAGGATGAA	1180
360	GlyHis	HisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAla	379
1181	GGGC	ATCATGTGTCAAAATTCAGGACAGAGCTCTGTGTAGTGGCTGCCACCTCTCATCT	1240
380	ArgTyr	IleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCys	399
1241	CCGT	ATCATTCAGGCTGCACATCTGCCGATTAAGGCTATTGACCTAGTTGATGAAGCTTGT	1300
400	AlaAsn	ValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLys	419
1301	GCAAT	GTGTAGATTCAGCTGTATGATCACTTCAGGAAATTCGAATCTTCGAGAGAG	1360

Qy	420	ArgMetGlnLeuGlnIleGlnLeuHiIsaLaLeuGluArgGluIysAspLysValaSerIys	439
Db	1361	AGAATTGAGCTGAGGTTGAACTTCAAGTCTTCGAGAGAAAGAAAGACAAGACTGACAA	1420
Qy	440	AlaArgLeuIleGlnValArgLysGlnLeuAspLeuArgAspLysLeuGlnProLeu	459
Db	1421	GCACGCTCATAGAAGTGAGAAAGAACTTGATGATTTGAGAGACAAATCCAAACCTTTG	1480
Qy	460	ThrMetLysTyrArgLysGlnLysGluArgIleAspGluIleArgGlnLysGlnLys	479
Db	1481	ATGATGAGGTACAAAGAAAGAGAAAGATAGATGAGCTCGCAGAGGTCAAGCAAAAG	1540
Qy	480	ArgGlnGlnLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAlaArgAla	499
Db	1541	CGGATGAGCTCATCATGCTTTTACAAAGAGCTGAAGAGAGATATGATCTGCGAGGGCA	1600
Qy	500	AlaAspLeuArgTyrGlyAlaIleGlnGlnValGlnLysAlaIleAlaGlnLeuGluGly	519
Db	1601	GCAGATCTGAGATATGGGCAATTCAAGAAAGTGCAAATGCATAGCAAAATCTTGAGAGT	1660
Qy	520	ThrSerSerGlnGluAsnValMetLeuThrGluAsnValGlyProGlnHiIleAlaGlu	539
Db	1661	ACCTCAGCT--GAAAGTACAAATGCTTAACAGAGACTGTGGGTCTCGATCAGATTGCGAA	1717
Qy	540	ValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGlnLysGluArg	559
Db	1718	GTGTGAGTGTGGACTGTGATATCCGGTCTCAAGGCTTGGTTCAGAATGAGAAAGAGAA	1777
Qy	560	LeuIleGlyLeuAlaAspArgLeuHiIsLysArgValValGlyGlnAsnGlnAlaValAsn	579
Db	1778	CTGATTTGGTCTTGGCGCATAGATTGCACAAAGAGTGGTGGGCAAGATCATGCAGTAGA	1837
Qy	580	AlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyValArgAlaGlnGlnProThr	599
Db	1838	GCTGTGCTGAGCCGTGTAAAGTCCAGAGCTGGTTAGGAAGGCCACAGCAACCACT	1897
Qy	600	GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGlnLeuAlaLysAlaLeu	619
Db	1898	GGTTCAITCTTTTCTTTGGGCGCAACTGGTGTGTGGAAGACAGAGCTCGCTAAAGGCTCT	1957
Qy	620	AlaGlnGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGluTyrMet	639
Db	1958	GCAGACAGACTCTTTTGATGATGATAAACTGATGATCAGATAGACATGTCGAGTACATG	2017
Qy	640	GluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHisGluGlu	659
Db	2018	GAAACAACACTCTGTTTCCGGCTGATTGGTGCTCCACAGGTTATGTTGGCATGATGAG	2077
Qy	660	GlyGlyGlnLeuThrGluAlaValArgArgArgProTyrCysValIleLeuPheAspGlu	679
Db	2078	GGAGGACAACCTACTGAAAGCTGTTAGAGAGCGGCTTACAGTGTGTGCTATTATGATGAA	2137
Qy	680	ValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGlyArg	699
Db	2138	GTTGAGAAAGCCATCTCTGCAGTGTTAATACATTGCTTCAAGTCTTCGATGATGGAAGG	2197
Qy	700	LeuThrAspGlyGlnGlyArgThrValAspPheArgHisSerValIleLeuMetThrSer	719
Db	2198	TTAACAGATGCTCAAGGCGGCACAGTGTGATTTCCACCAACTCCGTGATTAATGACTTCA	2257
Qy	720	AsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAla	739
Db	2258	AACTTGGGAGCAGAGTATCTGTGTCTGAGTAAATGGGCAAAATGATACCATGGAGACAGCT	2317
Qy	740	ArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeu	759
Db	2318	CGTGAATGGTTCATGCAGGAGGTGCGAAAGCAGTTTAAGCCGAGCTGCTGAATCGGCTG	2377
Qy	760	AspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeu	779
Db	2378	GATGAGATTGTTGTTTGTATCTCTGTCCCAAGCAGATGTGAGCGAAGTATGCCGTAC	2437

```

780 GlnMetIysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAsp 799
|||||
2438 CAGATGAAGACGCTGACCTACGCTGCTGAGAGGGTATTGCACTGGCGCTTACTGAG 2497
|||||
800 AlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIle 819
|||||
2498 GCAGCTCCAGATGCTACTACACAGAGAGTTATGACCGGCTTTATGTCGACAGACCTATT 2557
|||||
820 ArgArgTyrMetGluIysLysValValThrGluLeuSerLysMetValValArgGluGlu 839
|||||
2558 AGGATGGTGGAGAGAGAGTGGTACCGAGCTATCCAGATGCTCGTGAAGGAGGAG 2617
|||||
840 IleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGly-----AspLeuValTyr 857
|||||
2618 ATTGATGAGAACTCAACCGTTTACATAGATGCTGGGTCAGCGGGAAGATCTAAACCTAC 2677
|||||
858 ArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeu 876
|||||
2678 AGGTGGAGAGAAATGAGGCTTGTGAATGCTGCCACCGGCAAAATCTGATATATTG 2737
|||||
877 IleHisIleAlaAsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArg 896
|||||
2738 ATTCAAGCTCTTAATGTTCC---AGGATGATGCTGTCCAGCAGTCAAGAGATGAGG 2794
|||||
897 IleGluGluIleGluAspAspAsnGluGlu 907
|||||
2795 ATTGAAGAAATTGAAGATGACCAATGGAAGAT 2827
|||||

JLT 8
33840
US
INITIATION
Zea mays heat shock protein HSP101 (HSP101) mRNA, complete cds.
AF133840
AF133840.1 GI:4928487

Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 3084)
Young, F.E., King, J., Geisler-Lee, C.J., Tanguay, R.L., Caldwell, C.
and Gallie, D.R.
Developmental and thermal regulation of the maize heat shock
protein, HSP101
Plant Physiol. 127 (3), 777-791 (2001)
21563054
11706162
2 (bases 1 to 3084)
Young, T.E. and Gallie, D.R.
HSP101 Diversity in Monocot Species
Unpublished
3 (bases 1 to 3084)
Young, T.E. and Gallie, D.R.
Direct Submission
Submitted (09-MAR-1999) Biochemistry, University of California,
Watkins Drive, Riverside, CA 92521, USA
Location/Qualifiers
1..3084
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
1..3084
/gene="HSP101"
145..2883
/gene="HSP101"
/note="101 kDa protein"
/codon_start=1
/product="heat shock protein HSP101"
/protein_id="AAD33606.1"
/db_xref="GI:4928488"
/translation="MNPNFTHKTEAIVGAHEIAVEAGHAQLTPLHLAVALADKGG
ILROAITGASGDGAGAGSFERVLNLSLKLPSPDPDSVFASITALIKVIRRAQSAQ

```

```

KRGDSHIAVDOLLIGLLEDSQISDLCKEAGVSAARVAEAELEKRGGERVESASGD
TNFQALTKYRDLEVEQACKLDPIGRDEIRRVRIILSRRTKNPNVLIGEPGVGTAV
VEGLAQRNVRGDSNLDVRLIALDMGALVAGAKYRGEFEERLKAIVLEKEBAEKG
ILFDEIHLVAGRTEGSDAANLFKPMALGQRLCIGATTLEBYRYKVEKDAFER
RFOQVFVAEPSVPDVTSLRGLKEKEGHGVRIQDRALVVAALSAARYIMGRHLPOK
ALVDVEACANVRVQDSQPEIENDLERKIQLEVELHALKEKDKASKARLIEVRKE
LDDLKILQPLTMKTRKEKRIEIRKLQKREBLQFLQEAERMDLARVADLKYGA
LOEDDAALSKLESETGENMLTETVGPQIAEVSRSVMTGIPVIRLQGNKRLVLGLAD
RLHQVVOQTEAVASAVAEVARSAGLRPOOPTGSLFGLPTGVGKTELAKALAEQL
PDENLVRIDIMSEMEQHSARLIGAPPVGHGEGQLTEQVRRRYSVILFDEKE
KARVAFNTLLQVLDGRLTQGRTRVFNVTIIMTSLNLAELHLAGMVGNSKMA
RDLVMEVRRHERPELNLRIEIVFPLSHEOLKVKARLQMKVAVLAERGIATLAV
TDAALDIYSIDPYGCAPIRIRWIKRVVYVTSKMLIQEIDENCTVYIDAAPGKD
ELIVYKDRSGSLVNAETGKSDILLIQVNSSTRSDAAQAVKMKRIMEDEDEGMDEE"
BASE COUNT 670 a 874 c 1057 g 483 t
ORIGIN

Alignment Scores:
Pred. No.: 8,25e-167 Length: 3084
Score: 3883.50 Matches: 772
Percent Similarity: 92.88% Conservative: 76
Best Local Similarity: 84.56% Mismatches: 58
Query Match: 85.35% Indels: 7
Gaps: 6
DB: 8

US-09-812-350-17 (1-911) x AF133840 (1-3084)

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
DB 145 ATGAATCCGGACAACTTCAACCCACAAGACGAAGGCGCATCGTGGGGCGGCAGAGAT 204
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
DB 205 GCGGTGGAGCGCGGCACGCGCAGTCTACCGCGCTGCACCTGCGCGCGAGTGTGCTGCTCG 264
QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAla---GlyGlyGluAsnAlaAla 59
DB 265 GACAGGGCGGCATCTCGCGGACGAGCATACCGGGGGGTGCGGGGGCGCAGGACGCGCC 324
QY 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
DB 325 GGGGACTCTGTCGAGCGCGTGTGAACAACTCGCTCAAGAAGCTCGCTCGCGTCCGCG 384
QY 79 ProProAspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAla 98
DB 385 CGCGCGGACTCTCGTCCGCGCTCCAGCGCAGTCAAGAGTCAATCCCGCGCGCGCAGTCC 444
QY 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
DB 445 GCGCAGAAAAACGCGGGGACTCTCACTCGCGCTGCACCGAGTGTGCTCGCGCTGCTC 504
QY 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
DB 505 GAGGACTCGCAGATCTCCGACTCCCTCAAGAGGCGCGCGTGTGCGCGCGCGCGTGGCG 564
QY 139 SerGluValGluLysLeuArgGlyLysGlyLysLysValGluSerAlaSerGlyAsp 158
DB 565 GCGGAGCTTGAGAGCTCCGCGCGGGAGGCGCGCGCGTGGAGTCCGCGTGGCGGGGAC 624
QY 159 ThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeu 178
DB 625 ACCAATTCAGGCGCTCAAGACATACGCGCGGACCTGTTGAGCAGCGCGGAGCTG 684
QY 179 AspProValIleGlyArgAspGluLysLeuArgValValArgIleLeuSerArgArg 198
DB 685 GACCCCGTTCATCGCGCGGACGAGGAGATCCCGCGCTGCTGCGCATCTCTTCGCGCGCG 744
QY 199 ThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGlu 218
DB 745 ACTAAGAACAAACCCCTCTCTCAATCGCGGAGCGCGCGTGGAGAGCGCGCTGTTGGAG 804
QY 219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238

```

599 ThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLeuValThrGlyLeuAlaValAla 618
1942 ACTGGCTCGTTCTGTGTTCTGGGTCCGACTGGCGGGGAAACTGAGCTAGCCAGGCC 2001
619 LeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgLleAspMetSerGluTyr 638
2002 CTAGCCGAACAGCTGTTTCAGACGACGAACTTCTTGTCCGATCGACATGTCGGAGTAC 2061
639 MetGluGlnHisValSerValSerArgLeuLeuGlyAlaProProGlyTyrValGlyHisGlu 658
2062 ATGGAGCAGCACTCGGTTGCTCGCTCATCGAGCACCGCTCGCTAGCTCGCGCATGAA 2121
659 GluGlyGlnLeuThrGluAlaValArgArgArgProTyrCysValLleLeuPheAsp 678
2122 GAGGGTGGCAGCTGACTCAACAAGTGAAGAGGCGGTACAGCGTGATCCTGTTCCAC 2181
679 GluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGly 698
2182 GAGTTCAGAAAGGCCCATGTCCGCGGTTCACACCTGCTCCAGGCTCTCGACGACGCG 2241
699 ArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValLleIleMetThr 718
2242 AGGTTGACGATGGCAAGCAGCAGCGGTGACTTCAGAAACACCGTGATCATCATGACA 2301
719 SerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluVal 738
2302 TCGAACCTCGCGCCGAGCACCTCTCTCGTGGGTGGTGGCGAGAACTCCATGAAGGTC 2361
739 AlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArg 758
2362 GCTCGCATCTGGTTCATCGAGAGGTGAGAGGCATCTCCGACCCGAGCTCTGACCGT 2421
759 LeuAspGluLleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArg 778
2422 CTCGACGAGATCGTGATCTTCGATCCTCTGTCCACGAGCAGCTGAGGAAGTCCGTCGC 2481
779 LeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThr 798
2482 CTTCAGATGAAGATGTGGCGCTCTTGCAGAAAGGGCATCGCTCTGGCTGTGACC 2541
799 AspAlaAlaLeuAspTyrLleLeuAlaGluSerTyrAspProValTyrGlyAlaArgPro 818
2542 GACGCCGATTTGACATCATCTGTCTCTTACGATCCGGTGTACGGCGCGCGGCCA 2601
819 IleArgArgTrpMetGluLysLysValValThrGluLeuSerLysMetValValArgGlu 838
2602 ATCAGAGGTGGATCGAGAGAGGGTGTGACGACGCTGTGCAAGATGCTGATCCAGGAG 2661
839 GluLeuAspGluAsnSerThrValTyrLleAspAlaGlyAlaGly-----AspLeuVal 856
2662 GAGATCCGACGAGAACTGCACGGCTCTATCATCGACGCGCGCCGCAAGAGCAGCTGTGCT 2721
857 TyrArgValGlu----SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspVal 875
2722 TACAGGGTGGACCGAGCGCGGTGTGGTGAACGCTGAGACGGGATGAGTCCGACATC 2781
876 LeuLleHisLleAlaAsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMet 895
2782 CTGATCCAGGTCCTCCCAACAGCTCCACCGAGGCGACGCTCGCAGCGCTCAAGAAGATG 2841
896 ArgLle---GluGluLleGluAspAspAsnGluGlu 907
2842 AGGATCATGAGGAGGACGACGAGCGGTGTGGAACGCTGTGGAACGCGAGG 2880

RESULT 9
AF332981
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

AF332981 3148 bp mRNA linear PLN 02-JAN-2002
(HSP101) mRNA complete cds.

AF332981
AF332981.1 GI:18033449

Orvza sativa (japonica cultivar-group)

ORGANISM *Oryza sativa* (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoideae; Oryza.
 1 (Bases 1 to 3148)
 Agarwal, M., Sahi, C., Young, T., Agarwal, S., Mitra, S., Ganesan, K.,
 Agarwal-Katayar, S., Gallie, D.R. and Grover, A.
 Isolation and Analysis of HSP101 from Rice
 Unpublished
 2 (Bases 1 to 3148)
 Young, T.E. and Gallie, D.R.
 Direct Submission
 ITIE
 Submitted (02-JAN-2001) Biochemistry, University of California,
 Boyce Hall, Riverside, CA 92521-0129, USA
 Location/Qualifiers
 1..3148
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultiivar="Nipponbare"
 /db_xref="taxon:39947"
 1..3148
 /gene="HSP101"
 135..2873
 /gene="HSP101"
 /codon_start=1
 /product="heat shock protein"
 /protein_id="AA157165.1"
 /db_xref="GI:18033450"
 /translation="MNPNDFHKTNEALVAHEIASBAGHAQLTPHLVAALADKGG
 ILRAISQAGDAGAPDSFRRVSVGALKLPSPDPSPASTALIKVIRRAQSAQ
 KRGDSLAVDQLLGLLEDSLIDCLKEAGVSAARVALEKLRGEGEGRVSSASD
 TNFQALTYGRDIAVEQAGLDPVTGRDEIRRVVRIISRLTKNNPVLIGPGVGKTA
 ILFDEHLVLGAGETSGSDANLEKMLARGOLRCITGATTLVEYRKYVEKDAEPK
 RFQVFAEVPDITISLRGLKKEYGHHGVRIQDRALVVAQLSARYIMGHELPK
 AIDLDEACANRVOLDSQPEIDNLERIKRLEVEHHALEKGRKASKARLVKKE
 LDDKDLQPLTKRYKREKIDIRKLQREBELQFTLOEARRMDLARVADLUKGA
 LQEDIVAKLESETGETNMTETVGPQIAEVVSRMTGIPVTRIGQNDKERLVGLAD
 RLHORVQAGAVASAVAEVLRAGLRPOPTGSGFLFGPTGVGKTELAKALAEOL
 PDENLLVRIDMEVMEHSHVAELIGAPGVYGHGEGQLTEQVRRPYSVILPDEVE
 KAHVAFNTLLQVLDLGRDGTGQGTVDPRNTVIMTSLNGLAEHLLAGVGNKSMKVA
 RDLVMEVRRHFRHPLNRLDEIVDFLSHEQLRKVRLQMKDVAVLRAEVALVAV
 TDAALDVLSISYDPVGYGATIRWIKRVVTLQSKMLIOEIDENCTVYIDAAPHKD
 ELAVRDNRRGLVNAETGQKSDIITQVNGAATGSDAAQAVKQKRIWEDDEGMDER"
 644 a 931 c 1087 g 486 t
 E COUNT
 GIN

gment Scores:
 d. No.: 7,34e-165 Length: 3148
 re: 3840.50 Matches: 765
 cent Similarity: 92.00% Conservative: 75
 t Local Similarity: 83.79% Mismatches: 66
 ry Match: 84.41% Indels: 7
 Gaps: 5

09-812-350-17 (1-911) x AF332981 (1-3148)
 1 MetAnProGluLysPheThrHisLysThrAnGluThrIleAlaThrAlaHisGluLeu 20
 135 ATGAATCCGGAACATTCACGACAAAGACCAACAGAGCGCTCGTGGCGGCGCACGAGATA 194
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 195 GCGTCGAGCGGGCACGCGAGCTGACCGCTGCACCTGCTGGCGCGCTGGCGGCG 254
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAla--- 59
 255 GACAAGCGCGGATACCTCCGCGAGCGGATCTCGACGCGCTCCGCGGACGCGCGCG 314
 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 315 CCGGACTCGTTCGAGCGCGTGGTGGAGCGGCGCTCAAGAAAGCTGCCCTCGCAGTCGCCG 374

79 ProProAspAspIleProAlaSerSerSerSerLeuIleLysValIleArgArgAlaGlnAla 98
 375 CCGCGCGGACAGCGTCCGCGCTCCACGCGCTGATCAAGGTTCATCCGCGCGCGAGTTC 434
 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 435 GCGCAGAAAGCGCGGAGCTCCCACTTCGCGCGCGACCACTGCTGCTGGGCTCTCTG 494
 119 GluAspSerGlnIleArgAspLeuLeuAsnGlnValGlyValAlaThrAlaArgValLys 138
 495 GAGGACTCTCTAATCTCCGACTCCCTCAAGAGGCGCGCTGCTCCGCGCGCGGTGGCG 554
 139 SerGluValGlnLysLeuArgGlyLysGluLysLysValGluSerAlaSerGlyAsp 158
 555 GCGGAGCTCGAAGCTCCGCGCGGAGGCGCGCAAGGTGGAGTCCGCGCTCCGCGCGAC 614
 159 ThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGlnAlaGlyLysLeu 178
 615 ACCAATCTCCAGCGCTCAAGACCTACGCGCGCGACCTGCTCGAGCAGCGCGGTAACTC 674
 179 AspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArg 198
 675 GACCCGCTCATTTGGCGCGACGAGGAGATCCGCGCGCTGCTCGCATCTCTGCGCGCGC 734
 199 ThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValGlu 218
 735 ACAGAAACAAACCCCTACTCATCGCGAGCGCGCGCTCGGCAAGACGCGCTCGTGGAG 794
 219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
 795 GGCCTCGCGAGCGCATCTGCGCGGAGACGTGCCAGCAACCTGCTCGACGTCCGCTC 854
 239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGlu 258
 855 ATCGCGCTCGACATGGCGCGCTCTGCGCGCGCGCAAGTACCGCGCGCGAGTTCGAAGA 914
 259 ArgLeuLysSerValLysGluValGluAlaGluAspAlaGlyLysValIleLeuPhe 278
 915 CGGCTCAAGCGCGCTCAAGGAGGTGGAGGCGCGCGAGGTGATCTCTTCATC 974
 279 AspGluLysLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsn 298
 975 GAGCAGATACCTCTGCTCGCGCGCGGAGGACGAGGCGGTCCATGACGCGCGCAC 1034
 299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
 1035 CTGTTCAAGCCAATGCTCGCGAGCGCGAGCTCCGCTGCTCGCGCGCACGCGCTGGAG 1094
 319 GluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgPheGlnGlnValTyr 338
 1095 GAGTACCGCAAGTACGTCGAGAAGGACGCGCGCTTCGAGAGGCGGTTCAGACAGGTGTC 1154
 339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlyLysTyr 358
 1155 GTCCGCGAGCGAGCGCTCCCGGACCAATCAGCATCTCGAGGCGCTCAAGGAGAGTAC 1214
 359 GluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378
 1215 GAGGCGCACCAACGCTGAGGATCCAGGACCGCGCTCGTCTCGCGCGCGAGCTCTCG 1274
 379 AlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
 1275 GCGAGGTACATCATGGGACCGCCACCTGCTGCAAAAGCATGACCTCTGCTCGAGAGGCC 1334
 399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluIleAspAsnLeuGluArg 418
 1335 TCGCGCAACGCTGAGGTGAGCTGGACGACGACGAGGAGATCGACAACCTTCGAGAGG 1394
 419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSer 438
 1395 AAGAGGATTCAGCTGGAGGTGGAGCACCGCTGCGCGCGGAGGAGGAGGAGGAGGAGC 1454
 439 LysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnPro 458


```

1455 AAGGCCCGCGCTTGTGAGGTCAAGAAAGAGCTGGAGATCTGAGGAGCAAGCTCGAGCCA 1514
459 LeuThrMetLysTyrArgLysGluLysGluA7gIleAspGluIleArgLysGln 478
1515 TTGACGATGAAGTACAGGAGGAGGAGGAGGAGGATCGACGAGATCAGGAAGCTGAAGCAG 1574
479 LysArgGluGluMetPheSerLeuGlnGluA7gIleAspGluIleArgLysGln 498
1575 CGCCGCGAGGAGCTCGATTCAGGTTCAGGTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1634
499 AlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGlu 518
1635 GTCCGCGAGCTCAAGTACGCGCGCTCCAGAGATCGAGTCCGATCGGAGGAGTCCGAG 1694
519 GlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAla 538
1695 --AGCGAGACGGGGGAGAACCTGATGCTCAGCGAGCCGTCGCGCCGCGAGCAGATCGCC 1751
539 GluValValSerArgThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGlu 558
1752 GAGTGTGAGCGCTGGACCGGATCCCGGTGACCCGCTCGCGGCTCGCGGAGGAGGAGGAG 1811
559 ArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaVal 578
1812 AGGCTGTGCGCTCGCGGAGCTGCAGAGGAGTGTGTCGCGGAGGAGGAGGAGGAGGAGGAG 1871
579 AsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnPro 598
1872 AGCGCGTGTGCGGAGGAGGAGTGTGAGTGTGCGCGCGGAGTGTGCGCGGAGGAGGAGGAG 1931
599 ThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaVal 618
1932 ACCGGTGTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1991
619 LeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGluTyr 638
1992 CTGCGCGAGCAGCTGTGCGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2051
639 MetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGlu 658
2052 ATGAGCAGCAGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2111
659 GluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAsp 678
2112 GAAGTGGGAGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2171
679 GluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGly 698
2172 GAGTTCGAGAGGCGCAGCTCGCGGTGTTCAACACTCTGCTCCAGGTTCTCGAGCAGCGG 2231
699 ArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleLeuMetThr 718
2232 AGGCTCACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2291
719 SerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluVal 738
2292 TCCACCTCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2351
739 AlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArg 758
2352 GCCCGTGCACCTTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2411
759 LeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArg 778
2412 CTCACGAGATTTGATATTCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2471
779 LeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThr 798
2472 CTCAGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2531
799 AspAlaAlaLeuAspTyrIleLeuAlaIleSerTyrAspProValTyrGlyAlaArgPro 818

```

```

2532 GAGCGCGCGCTGACGATCTGTCTGTCTATGATCCGGTTTATGCGCGAGCGCG 2591
819 IleArgArgTrpMetGluLysLysValValThrGluLeuSerLysMetValValArgGlu 838
2592 ATCAGGAGGTGGATCGAGAGAGGGTGTGACCGATGTCGCAAGATGTCGATCCAGGAG 2651
839 GluIleAspGluAsnSerThrValTyrIleAspAlaGly-----AlaGlyAspLeuVal 856
2652 GAGATCGACGAGAACTGCACGGTGTATCATCGACGCGCGCGCACCAAGGACGAGCTGGCC 2711
857 TyrArgValGluSer---GlyGlyLeuValAspAlaSerThrGlyLysLysSerAspVal 875
2712 TACAGGCTGACCAACCGCGCGGGTGTGTGAACCGCGAGACGGGGGAGAGTCCGACATC 2771
876 LeuIleHisIleAlaAsnGlyProLysArg---SerAspAlaAlaGlnAlaValLysLys 894
2772 CTATCCAGGTGCGCAACGGCGCGCGCCACCGGAGGCGACCGCGCGCGGCGGTGAGAG 2831
895 MetArgIleGluGluIleGluAspAspAspAspGluGlu 907
2832 ATGAGGATCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2870

RESULT 10
AF097363 3107 bp mRNA linear PLN 14-MAY-2002
LOCUS Triticum aestivum heat shock protein 101 (Hsp101b) mRNA, complete
DEFINITION cds.
ACCESSION AF097363
VERSION AF097363.1 GI:4558483
KEYWORDS Triticum aestivum (bread wheat)
SOURCE Triticum aestivum
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 3107)
AUTHORS Campbell,J.L., Klueva,N.Y., Zheng,H.G., Nieto-Sotelo,J., Ho,T.D.
and Nguyen,H.T.
TITLE Cloning of new members of heat shock protein HSP101 gene family in
wheat (Triticum aestivum (L.) Moench) inducible by heat,
dehydration, and ABA(1)
JOURNAL Biochim. Biophys. Acta 1517 (2), 270-277 (2001)
MEDLINE 21240210
PUBMED 11342108
REFERENCE 2 (bases 1 to 3107)
AUTHORS Campbell,J.L., Klueva,N.Y. and Nguyen,H.T.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1998) Plant and Soil Science, Texas Tech
University, Mail Stop 2122, Lubbock, TX 79409, USA
FEATURES
Location/Qualifiers
1..3107
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Mustang"
/db_xref="taxon:4565"
/dev_stages="10 day old seedling"
/notes="treated 2 hrs. at 37 degrees C."
1..3107
/gene="Hsp101b"
112..2868
/gene="Hsp101b"
/notes="ClpB family member"
/codon_start=1
/product="heat shock protein 101"
/protein_id="AAD22629.1"
/db_xref="GI:4558484"
/translation="MNPDNFTKTNELVRAHEAASEAGHAQITPLHLAALAAADKSG
ILRQAVAGSAGNSAGDSFVSLGALKLPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
KKRGSHLAVQLLMGLLEADQIADCKEAGSRSVRAELEKLRGNDGRKRVESAG
DTNFKALYGRDLVEVAGKLDPIGRDEIRVRVIRLSRRTKNNPVLIGPGVGKTA
VVEGLAQRVVRGDPVSNLLDVLVALDMGLVAGKVRGSEFEERLKVAVEKEAEKG
VIFIDEHLVLGAGRTGSDAANLKPMLARGQLRCIGATTLEEVRKVKVEKDAFE
RRFQVTVAEPSVADTISILKGLKEVGHGVRIQDRAIVVAQLSARYIMGRHLDP

```

KAIDLVDACANVRVOLDSPBEIDNLRKRIQLEVEHLEKEKDKASKARLVEVRK
 ELDLRLKQLPQTKYRKERIDEIRKIKORRELOFTQAEARRMDLARVADLVKYG
 ALQEDIAALKEGTGENMLTETVGPBQAEVVSRTGTPTVRLQNDKERLVGMA
 DRLHTRVVGQTEAVNAVAEIRAGLRGQQPTGSGFLGPTGSGVGTAKALABQ
 LFDENLIVRIDMSYEQHSVARELIGAPPYVGHHEGGQLTEQVRRRPSYVLFDE
 EKAVAVENTLLODDRLTDGQORTVDFRNTVIMTSLGAHEHLLAGMVGKNSKV
 ARDLVMOVRHRFRPELNLDEIVFPDPLSHEOLRVARLQKQVAVLRAERGVALA
 VTDAALDLSLAVDPVGPASPIRWIEKRVVLTOLSKWMLQIEIDENCTVYIDAANKO
 ELAYRDRSGGLVNAETQORSDDLLOVFNALGSGKGGEAAKAVKQWRVMDGDDEGM
 DEDV"

SEQUENCE COUNT 605 a 954 c 1116 g 432 t

Alignment Scores:

id. No.: 3107
 Length: 3107
 Matches: 760
 Conservative: 82
 Mismatches: 66
 Indels: 12
 Gaps: 7

us-09-812-350-17 (1-911) x AF097363 (1-3107)

1 MetAspProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 112 ATGACCCCGGACAACTTCACGCACACAGACCAAGAGCGCTGCTGCGCGCGCACGAGCGG 171
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisIleAlaGlyAlaLeuIleSer 40
 172 GCGTCGAGCGCGCGCACGCGAGATCACGCGCTGCACCTGCGCGCGCGCGCTGCGCGCG 231
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAla--- 59
 232 GACAACTCGGGGATCTCGCGACGCGCGCTGCGCGCGCGCTGCGCGCGCGCGCGCGCG 291
 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 292 GCGAGCTGCTGCGAGCGCTGCTCTCCGCGCGCTCAAGAACTGCGCGCGCGCGCGCGCG 351
 79 ProProAspIleProAlaSerSerSerSerLeuIleLysValIleArgArgAlaGlnAla 98
 352 CGCGCGGACTCGCGCGCGCTCCACGCGCGCTCATCAAGCGCATCGCGCGCGCGCGCGCT 411
 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 412 GCGCAAGAAGCGCGGACTCCCACTTCGCGCTGCACACAGCTGCTCATGCGCGCTCTCT 471
 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
 472 GAGGACGCGCAGATCGCGACTGCTCAAGAGGCGCGCGCTGCTCGCGCTGCGCGCTGCG 531
 139 SerGluValGlnLysLeuArgGlyLysGlu---GlyLysValGluSerAlaSerGly 157
 532 GCGAGCTCGAAGACTCGCGCGCGGGAACCTCGCGCAAGTTCGAGTTCGCGCTCGCGCG 591
 158 AspThrAsnPheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAlaGlyLys 177
 592 GACACCAACTTCAGCGGCTCAAGACGACGCGCGGACTCGCTCGAGGTGCGCGCGCAAG 651
 178 LeuAspProValIleGlyArgAspGluIleArgArgValValArgIleLeuSerArg 197
 652 CTCGACCCGCTATCGCGCGCGAGAGATCGCGCGCTGCTGCGCATCTCTCTCACGC 711
 198 ArgThrLysAsnAsnProValLeuIleGlyLysProGlyValGlyLysThrAlaValVal 217
 712 CGCACCAAGAACACCCCGCTCTCATCGCGAGCGCGCGCTGCGCGCAAGCGCGCGCTG 771
 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 772 GAGGCGCTCGCGAGCGCGCTGCTGCGCGCGGAGCGTGCACGACCACTGCTCGAGCTCG 831
 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrglyGluPheGlu 257

832 CTGCTCGCGCTCGACATGCGCGCGCTGCTGCGCGCGCGCAAGTACCGCGCGGAGTTCGAG 891
 258 GluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPhe 277
 892 GAGCGCTCAAGCGCTGCTCAAGGAGGTGGAGGAGCGCGAGGCGAGGTCTCTCTCTTC 951
 278 IleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAla 297
 952 ATCGACGAGATACACCTGCTGCTGCGCGCGCGGAGCGAGGCGGTCTCATGAGCGCGCC 1011
 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeu 317
 1012 AACCTGTTCAGCGCGATGCTTCGCGAGGCGCGAGCTCCGCTGTCATCGCGCGCACAGCTG 1071
 318 GluGluTyrglyTyrglyValGluLysAspAlaAlaPheGluArgArgPheGlnVal 337
 1072 GAGGAGTACCGCAAGTACGTCGAGAGAGGAGCGCGCTTCGAGCGCGGTTCACAGAGGTG 1131
 338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLysGluLys 357
 1132 TAGTGGCGCGAGCGCGCTGCGCGCACCATCAGCATCTCTGCGAGGCGCTCAAGGAGAG 1191
 358 TyrGlyGlyHisHisGlyValArgGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377
 1192 TACGAGGCGGACCAACCGCGCTGCGCATCCAGGACCGCGCGCTGCTGCGCGCGCGAGCTC 1251
 378 SerAlaArgTyrglyThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 1252 TCGCGAGAGTACATCATGCGCGCGCGCGCTGCGCGCACAGGCGCATCGACTGCTGGTGG 1311
 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluIleLeuAspLeuGlu 417
 1312 GCGTGCACCAACGTGAGGTGCGAGTGCAGACGCGCGCGAGGAGATCGCAACCTGGAG 1371
 418 ArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAla 437
 1372 AGGAAGCGGATTCAGCTGGAGGTGAGCTGACGCGCTGGAGAGGAGAGGAGGAGGAGG 1431
 438 SerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGln 457
 1432 AGCAAGCGCGCTGCTGGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1491
 458 ProLeuThrMetLysTyrglyArgLysGluLysGluArgIleAspGluIleArgArgLeuLys 477
 1492 CCGCTGACGATGAAGTACAGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1551
 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrglyAspLeuAla 497
 1552 CAGCGCGCGAGGAGCTGCGAGTTCACGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1611
 498 ArgAlaAlaAspLeuArgTyrglyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 517
 1612 CGCGTTCGCGACCTCAAGTACGCGCGCTCCAGGAGATCGACGCGCGCATCGCGCAAGCTG 1671
 518 GluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
 1672 GAGGCGGAGAGCGCG---GAGAACTGATGCTCCAGAGACCGCTGCGCGCGCGCGAGCATC 1728
 538 AlaGluValValSerArgTyrglyThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
 1729 GCGGAGGTGGTGGAGCGGTGACCGGTGACCGCATCCCGGTGACCGCGCTCGGCGAGAGCA 1788
 558 GluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
 1789 GAGCGGCTCGTCGCGCATCGCGGACAGCGCTGCACACGAGGCGGTGGTGGCGGAGAGGCG 1848
 578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGln 597
 1849 GTGAACGCGCTCGCGGAGCGGTGCTGAGGTCCAGGCGCGGCTTGGCGCGCGCGAGCAG 1908
 598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLys 617
 1909 CCCACCGGCTGCTCTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1968

618 AlaLeuAlaGluGlnLeuPheAspGluAsnLeuLeuValAlaGleAspMetSerGlu 637
 1969 GCGCTCGCGGAGACCTGTTGACGACGAGAACCTGCTGTCGCGCATCGATGTCGAG 2028
 638 TyrMetGluGlnHisSerValSerArgLeuGlyAlaProProGlyTyrValGlyHis 657
 2029 TACATGGAGCAGCACTCGCTCGCGCGCTCATCGGAGCGCCACCTGGCTATGTCGGGCAC 2088
 658 GluGluGlyGlyGlnLeuThrGluAlaValAlaArgArgProTyrCysValIleuPhe 677
 2089 GAGGAGGCGCGGACGCTGACGAGCAGGTGAGGAGGCGGTACAGCGTATCTGTTTC 2148
 678 AspGluValGluLeuAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp 697
 2149 GACGAGGTGGAGAGCGGACGCTGGCGGTGTTCAACAGCTCTCCAGGTGCTGGAGCAG 2208
 698 GlyArgLeuThrAspGlyGlnGlyArgThrValAlaPheArgAsnSerValIleuMet 717
 2209 GCGCGGCTGACCGAGCGGCAAGCGGAGGAGCGGTGAGCTTCGGAACACCGGTGATCATCATG 2268
 718 ThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGlu 737
 2269 ACTTCACCTCGCGCGGCGGAGCACCTCTCGCGGATGCTGCGGAGCAAGCACTCCATGAG 2328
 738 ValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsn 757
 2329 GTCGCGCGGACCTGCTCATGATGAGGAGGTGAGGAGCACTTCAGGCGGAGCTGCTGAAC 2388
 758 ArgLeuAspGluLeuValPheAspProLeuSerHisAspGluLeuArgLysValAla 777
 2389 CGCTGAGCAGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2448
 778 ArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaVal 797
 2449 CGCTGAGATGAGGAGCGGTGCGGCTGCGGCTGCGGAGCGGCGGTGCGCTGCGCGTC 2508
 798 ThrAspAlaAlaLeuAspTyrIleuAlaGluSerTyrAspProValTyrGlyAlaArg 817
 2509 ACCGAGCTGCGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2568
 818 ProIleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArg 837
 2569 CCGATCAGGAGGTGGATCAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2628
 838 GluGluLeuAspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAsp---LeuVal 856
 2629 GAGGATGACGAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2688
 857 TyrArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspVal 875
 2689 TACCGGTGAGCGGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2748
 876 LeuIleHisIleAlaAsnGlyPro-----LysArgSerAspAlaAlaGlnAla 891
 2749 CTATCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2808
 892 ValLysLysMetArgIleGluGluLeuGluAspAspAsnGluLeuMetIleGluAsp 911
 2809 GTGAGNAGATCGGGTC-----ATGAGAGCGGGGAGGAGCGGAGCGGAGCATGAGAGAT 2862
 AFI74433 3058 bp mRNA linear PLN 14-MAY-2002
 Triticum aestivum heat shock protein 101 (HSP101c) mRNA, complete cds.
 AFI74433 GI:6013195
 Triticum aestivum (bread wheat)
 Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Triticum.
 1 (bases 1 to 3058)
 Campbell, J.L., Klueva, N.Y., Zheng, H.G., Nieto-Sotelo, J., Ho, T.D. and Nguyen, H.T.
 Cloning of new members of heat shock protein HSP101 gene family in wheat (Triticum aestivum (L.) Moench) inducible by heat, dehydration, and ABA(1)
 Biochim. Biophys. Acta 1517 (2), 270-277 (2001)
 21240210
 MEDLINE
 11342108
 PUBMED
 2 (bases 1 to 3058)
 Klueva, N.Y., Campbell, J.L., Zheng, H. and Nguyen, H.T.
 Direct Submission
 Submitted (02-AUG-1999) Plant and Soil Sciences, Texas Tech University, Mail Stop 2122, Lubbock, TX 79409, USA
 JOURNAL
 Location/Qualifiers
 1. 3058
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultiivar="Mustang"
 /db_xref="taxon:4565"
 1. 3058
 /gene="HSP101c"
 94. 2835
 /gene="HSP101c"
 /note="C1pB family member"
 /codon_start=1
 /product="heat shock protein 101"
 /protein_id="AAF01280.1"
 /db_xref="GI:6013195"
 /translation="MNPDKFTHTNEALAAAHMASEAGHAQITPLHAAALAAADRSG ILQTAHAGSGNDAAARSFVASALKLPSSQPPDTPASTSLVKAVRAQSAQ KSRGSHLAVDQIMGLLEDPOISDALXGANGISAARYKAIVEKLGDNRRVESAGD TNGQALYGRDVEVAGKLDVPIGRDEIRRVIRLSRTKNNPVLIGEPGVGKTAV VEGALVRVGGDVPVSNLDVRLVDMLGALVAGKYRGEFEELKAVLKEVEAEQV ILFDTEHVLGAGRTGGMDAANPKMLRQALQICIGATTLTEYRYKVEKDAFER RFOQVFAPSPVPTVSIILGLKKEVGHGVRIQDRALVIAQLSSRYIMGRHLPDK AILVDEACANVQLDSQPEIDNLNLRKRIOLEVELHALEKEDKASKARLVDRKE LDDRLKLPQMKYRKERKIDRISLQKORRELQTLQEAERMDLARVADLRGA LQEDRAIAKLEGTENLMTETVPDQAEVVSWRTGIPVTRLCQNEKARLIGLAD RLQRVVVGOYEVAVNGEAVIRAGLRPQPTGSGFLGPTGVGTRELAKALASQL FDNELVIRIDMSEYMEQSVARLIGAPPYGVGHEGQLTQVRRRPSYVILDFVE KAHVAVNTLLVDLDDGLTDGRTVDFRNTVIIMTSLGASHLAVAGVGNMKTAR DLYWQVRRFRPELLNLRLDEIVIFDLSHEQLRKVARLQMKDVAVLAERGVALVT DAALDVILISUSYDPVYGAERFIRMIKRIEIVTELSKMLIREIDENSTVIDAAPSKE LTGVDRKHGGLVNARTGHKSDILIQVPSGAVGGDAHAHVKKIMQDSVEVDMERE"

BASE COUNT 680 a 836 c 1025 g 517 t
 ORIGIN

Alignment Scores:
 Pred. No.: 4.62e-164 Length: 3058
 Score: 3822.50 Matches: 761
 Percent Similarity: 91.47% Conservative: 75
 Best Local Similarity: 83.26% Mismatches: 69
 Query Match: 84.01% Indels: 9
 Gaps: 6

US-09-812-350-17 (1-911) x AFI74433 (1-3058)

QY 1 MetaEnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 Db 94 ATGATCCCGACAGTTCACGACACCAATAGCGGCTCGCGCGCGCGCGCATG 153
 QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 Db 154 GCGTCGAGGCGCGCACCGGACCTCACCGCTGACCTCGCGCGCGCGCTCGCGCG 213
 QY 41 AspProThrGlyIlePheProGlnAlaIleSerAlaGlyGlyGluAsn-----Ala 58
 Db 214 GACAGTGGCGGACCTCTCGCGCAGGCCATGCCACCGCTCCCGCGGCGGCGCGCG 273
 QY 59 AlaGlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78

* are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 33927: contig of 33926 bp in length
 * 34026: gap of unknown length
 * 98487: contig of 64461 bp in length
 * 98488: gap of unknown length
 * 114278: contig of 15691 bp in length
 * 114279: gap of unknown length
 * 123438: contig of 9060 bp in length
 * 123439: gap of unknown length
 * 123539: contig of 23918 bp in length
 * 147457: gap of unknown length
 * 147557: contig of 14873 bp in length.

Location/Qualifiers

1. 162429
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="5"
 /clone="P0599F04"
 45748 a 34689 c 34184 g 47308 t 500 others

Alignment Scores:

d. No.: 2.63e-156 Length: 162429
 3691.00 Matches: 765
 81.00% Conservativity: 75
 73.77% Mismatches: 66
 81.12% Indels: 132
 2 Gaps: 9

us-09-812-350-17 (1-911) x AC132491 (1-162429)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 120111 ATGAATCCGAGCAACTTCACGCAACAGCAACAGGCGCTCGTGGCGGCGCAGCAGATA 120052
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 120051 GCGTCGAGGCGGGGACGCGCGAGCTGACGCGCTGCACTGTGTGGCGCGTGGCGGG 119992
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAla--- 59
 119991 GACAAGGGCGGATACCTCCGCGAGCGGATCTCGCAGCGCTCCGCGGCGGCGCGCGCG 119932
 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLeuProSerGlnSerPro 78
 119931 CCGGACTCGTTCGAGCGCGTGGAGCGGCGGCGCTCAAGAAGCTGCGCTCGCAGTCGCG 119872
 79 ProProAspIleProAlaSerSerLeuIleLysValIleArgArgAlaGlnAla 98
 119871 CCGCGGACAGCGTGGCGGCTCACGCGCTGATCAAGTCTATCCCGCGCGCAGTCC 119812
 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 119811 GCGCAGAAGAAGCGCGGACTCCACCTCGCGTGCAGCAGCTGTGCTGGCGCTCTCTG 119752
 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
 119751 GAGACTCGCTATCTCCGACTCGCTCAAGAGCGCGCGTGTCCCGCGCGGGTCCGC 119692
 139 SerGluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAsp 158
 119691 GCGGAGCTCGAAGACTCCGTGGCGGGGGGCGCAGAGTGGAGTCCGCGTCCGCGGAC 119632
 159 ThrAsnPheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAlaGlyLysLeu 178

Db 119631 ACCAACTTCAGGCGCTCAAGACTACGCGCGCGAGCTCGTCGCGAGCGCGCTAAAGCTC 119572
 Qy 179 AspProValIleGlyArgAspGluGluIleArgValValArgIleLeuSerArgArg 198
 Db 119571 GACCCCGTCAATGGCGCGAGAGATCCGCGCGTGTGTGCGATCTCTGTGCGCGCG 119512
 Qy 199 ThrLysAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGlu 218
 Db 119511 ACGAAGAACCAACCCGCTACTCATCGCGAGCGCGCGTCCGCAAGCGCGCTGTGGAG 119452
 Qy 219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
 Db 119451 GGCCTCGCGAGCGCATCTGTCGCGGGGACGTGCCAGCAACCTGTCTCGACGCTCCGCTC 119392
 Qy 239 IleSerLeuAspMetGlyAlaLeuValAlaGlyValLysThrArgGlyGluPheGluGlu 258
 Db 119391 ATCCGCTCGACATGGCGCGCTGTCCGCGCGGCGCAAGTACCGCGCGAGTTCGAGAGAA 119332
 Qy 259 ArgLeuLysSerValLeuLysGluValGluAspAlaGlyLysValIleLeuPheIle 278
 Db 119331 CGGCTCAAGCGCGTCTCAAGGAGGTGGAGGCGGAGGCAAGGTGATCTCTTCATC 119272
 Qy 279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluClySerMetAspAlaAlaAsn 298
 Db 119271 GACGAGATACACCTCTCTCGCGCGCGGAGGACGAGGCGCTCCATGGACGCGGCGAAC 119212
 Qy 299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
 Db 119211 CTGTTCAAGCCATCTCGCGGCGGCGCAGCTCGGTGCTATCGCGCGCACGCTGGAG 119152
 Qy 319 GluTyrrArgLysTyrrValGluLysAspAlaAlaPheGluArgArgPheGlnValTyrr 338
 Db 119151 GAGTACCGCAAGTACGTCCGAGAGGACGCGCGCTTCGAGAGGCGGTTCACGAGGTGTTTC 119092
 Qy 339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysTyrr 358
 Db 119091 GTCCGCGAGCGGAGCGTCCCGGACACCATCAGATCTCGGAGCGCTCAAGGAGAGTAC 119032
 Qy 359 GluGlyHisLysGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378
 Db 119031 GAGGGGACCAACGCTGTGAGATCCAGACCGCGCGCTCGTCTCGCGCGCAGCTCTCG 118972
 Qy 379 AlaArgTyrrIleThr----- 383
 Db 118971 GCGAGGTATCATCAT-GGGTAAGTTGCTCTCGCGCTCTTCCATTGCGCAACCTA 118913
 Qy 384 -----GlyArg 385
 Db 118912 CATGTCAAGCTCGAGTTCAATTTGTGCGCATTTCTGTCTCTGTAAATTTTCAGGACGC 118853
 Qy 386 HisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAlaAsnValArgValGln 405
 Db 118852 CACCTGCTGCAAAAGCGATCGACCTGTCGAGAGGCGCTCGCGAAGCGTGGAGGTGAG 118793
 Qy 406 LeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArgMetGlnLeuGluIle 425
 Db 118792 CTGACACCGCAGCGAGAGATCGAACCTCGAGAGGAGGATTCAGCTGGAGGTG 118733
 Qy 426 GluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAlaArgLeuIleGluVal 445
 Db 118732 GAGCACCGCGCTGGAG 118673
 Qy 446 ArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThrMetLysTyrrArgLys 465
 Db 118672 AAGAAAGAGCTGGACGATCTGAGGACAAAGCTGAGGCAATTCAGATGAAGTACAGAG 118613
 Qy 466 GluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArgGluGluLeuMetPhe 485
 Db 118612 GAGAAGGAGAGAGATCGACGAGATCAGAGAGCTGAAGCAGCGCGCGAGGAGCTCCAGTTC 118553
 Qy 486 SerLeuGlnGluAlaGluArgArgTyrrAspLeuAlaArgAlaAlaAspLeuArgTyrrGly 505
 Db 118552 ACGCTGACAGGAGCGCGAGCGCGGATGAGCTCTCGCGCGCTCGCGGACCTCAAGTACGCG 118493

Qy	808	uSerTyrAspPro		-----	812
Db	117415	GrCCCTATGATCGGTACGCGACAGCATCCA	ACTATCAGGTCTGAAGTTTTTCA	AGAAATT	117356
Qy	813	-----	-----ValTyrGlyA	816	
Db	117355	CTCGTCGAATTGATCTTA	CTCTCTGTGCACTCTCGGTACTACGTACAGGTTTATGGCG		117296
Qy	816	laArgProileAArgArgTirpMetGluLysLysValValThrGluLeuSerLysMetValV		836	
Db	117295	CGAGCCCGATCAGGAGGTGGATCGAGAAGAGGGTGGTGC	CGCAGCTCTCCAAGATGCTGA		117236
Qy	836	aIArgGluGluLeuAspGluAsnSerThrValTyrIleAspAlaGly	-----AlaGlyA	854	
Db	117235	TCCAGGAGGAGATCCAGCAGAACTGCACGGTGTACATCGACCGCGCGCCGACAAAGAGC			117176
Qy	854	spleuValTyrArgValGluSer	--GlyGlyLeuValAspAlaSerThrGlyLysLys	873	
Db	117175	AGCTGGCCCTACAGGTGTGAACAACCGCGGGGGTGGTGAACCGCGACCGGGGCAGAAGT			117116
Qy	873	erAspValLeuIleHisIleAlaAsnGlyProLysArg	---SerAspAlaAlaGlnAlaVal	892	
Db	117115	CGGACATCTCTATCCAGGTGCGCAACGCGCGCCGCCACCGGGAGCGACCGCGCAGCGCG			117056
Qy	892	aLysLysMetArgIleGluGluLeuAspAspAsnGluGlu	907		
Db	117055	TCAGAAGATGAGGATCATGAGGACCGAGGACCGCATGCGCAGAG			117009

```

RESULT 15
OSAS16025
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

OSAS16025
Oryza sativa hsp101 gene for heat shock protein 101, exons 1-5.
AJ316025
AJ316025.1 GI:22535405
heat shock protein 101; hsp101 gene.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1
Agarwal,M., Sahi,C., Young,T., Agarwal,S., Mitra,S., Ganeshan,K.,
Katiyar-Agarwal,S., Gallie,D.R. and Grover,A.
Isolation and analysis of HSP101 from rice
Unpublished
2 (bases 1 to 3109)
Agarwal,M.
Direct Submission
Submitted (28-AUG-2001) Agarwal M., Department of Plant Molecular
Biology, University of Delhi South Campus, Benito Juarez Road,
Dhaura Kuan, New Delhi, Delhi-110021, INDIA
Location/Qualifiers
1. .3109
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
3. .3108
/gene="hsp101"
join(3..1159,1258..2061,2150..2431,2516..2710,2808..3130)
/gene="hsp101"
/codon_start=1
/product="heat shock protein 101"
/protein_id="CAC87117.1"
/db_xref="GI:22535406"
/tranlation="MNPNFTHKTNELVAAHETASEAGHAQLTFLHVAALAADKGG
ILRQIAISAGDADGAPSPRVSVGALKKLPSQPPDPSVPASTALIKVIRRAQSAQ
KKRGSHLAVQLLIGLLESDLCLEAGVSAARVAELEKRGCEKRVKESAGD
TNFQALTYGRDVEQAGKLDPIVGRDEEIRRVVRLISRETKNPNVLIGPGVGKTAI
VSLGAKTVGRDVSFNLNDLRLIADMGLVAGKATGFEERLKAIVKEVEAEKSV
ILFDIEHLVLAGQETSGSDMAALFKPMLGARGQVLCIGATTLEETKRVYVEKDAFER
IFQOYFVAEYSPDPTISLIGKEGLKEHGVIRIQDRALVVAQLSARYIMGHRLDPK

```


LDDLRKLOPLTKYKKEKERIDEIRKLKORREELQFTLQEAERRMOLARVADLKVGA
 LQEDIVAKLSEETGENMLTETVGPQIAEVSRWTGIVPTRLQONDEERLVLGLAD
 RHQVRVGAQAVSAVAARVSRAGLRPOQPTGFLPLGPTGVGKTELAKALAEQL
 FDEENLLRIDMSYMEQHSVARLIGAPPGVYHGGGLTEQVRRRYPYVILYDEVE
 KAHVAFNTLLQVLDGRLTDGQRTVDPRNTVIMTSLNGAEHLLAGVMKSNKVA
 ROLVQEVRRHFRPELLNRLDEIVLFDPLSHEQLKVARLQMKDVAVRLAGRGVALAV
 TDAALDVLISLSDVPYVGARPIRRWIEKRVVTQLSKMLIOBEIDENCTVYIDAAPHKD
 ELAYVRDRRGLVNAETGQKSIDILLIQVNGAATGSDAQAQVKWRIIMEDEGMDDEE"
 <3..1159

exon /gene="hsp101"
 /number=1
 intron 1160..1257
 /gene="hsp101"
 exon /number=1
 1258..1261
 /gene="hsp101"
 /number=2
 intron 2062..2149
 /gene="hsp101"
 /number=2
 exon 2150..2431
 /gene="hsp101"
 /number=3
 intron 2432..2515
 /gene="hsp101"
 /number=3
 exon 2516..2710
 /gene="hsp101"
 /number=4
 intron 2711..2807
 /gene="hsp101"
 /number=4
 exon 2808..3108
 /gene="hsp101"
 /number=5

BE COUNT 617 a 951 c 1048 g 493 t
 IGIN

Alignment Scores:
 Seq. No.: 1,24e-156 Length: 3109
 Score: 3658.00 Matches: 760
 Percent Similarity: 80.71% Conservative: 77
 Local Similarity: 73.29% Mismatches: 69
 Match: 80.40% Indels: 132
 Gaps: 9

us-09-812-350-17 (1-911) x OSA316025 (1-3109)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 3 ATGAATCCGACAACTTCAGCACAAAGACCAACGAGGCGCTCTGTGGCGCGCACAGAGATA 62
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 63 GCCTCGAGCGCGGCGCACGCGCAGCTGACGCGCGCTGCACCTGTGTGGCGCGCTGCGCGG 122
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAla--- 59
 123 GACAAAGCGCGGATCTCCGCGAGGCGATCTCCAGCGCTCGCGCGGCGACGCGCGGCG 182
 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 183 CCGGACTCGTTGAGCGCGTGGTGAGCGGCGCTCAAGAAGCTGCGCTCGCAGTCCGCGG 242
 79 ProProAspIleProAlaSerSerSerLeuIleIleValIleArgAlaGlnAla 98
 243 CCGCGGACAGCGTGGCGGCTCATCAAGTCTATCCGCGCGCGCGCAGTCC 302
 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeu 118
 303 CGCGAGAGAGCGCGGAGCTCCACCTCGCGTGCACGAGCTGTGTGGCGCTCTG 362
 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138

363 GAGGACTCGCTATCTCCGACTCCCTCAAGAGCGCGCGTGTCCGCGCGCGGTGCGC 422
 139 SerGluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAsp 158
 423 GCCGAGCTCAGAAGCTCCGTGGCGGAGCGCGCAAGGTGGAGTCCGCGTCCGCGGAC 482
 159 ThrAsnPheGlnAlaLeuLysThrTyTGlyArgAspLeuValGluGlnAlaGlyLysLeu 178
 483 ACCAACTTCCAGGCGCTCAAGACTACGCGCGGAGCTCTGTGAGGAGCGCGTAAAGTTC 542
 179 AspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArg 198
 543 GACCCGCTCATTTGGCGCGACGAGAGATCCGCGCGCTCGTGGCATCTCTGTCCGCGCGC 602
 199 ThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValGlu 218
 603 ACGAAGAACCAACCCGCTACTCATCGCGGAGCGCGCGTTCGCGAAGACGCGCGTGGAG 662
 219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
 563 GGCCTCGCGCAGCGCATCTGTGCGGGGACGTGCCAGCAACCTGCTCGACGTCGCGCTC 722
 239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyArgGlyGluPheGluGlu 258
 723 ATCGCGCTCGACATGGCGCGCTCTGTCCGCGGCGCAAGTACCGCGCGAGTTCGAAGA 782
 259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
 783 CGGCTCAAGCGCGCTCAAGGAGGTGGAGGAGCGCGGAGGCTCCATGAGCGCGCGCAAC 842
 279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGlySerMetAspAlaAlaAsn 298
 843 GACGAGATACACCTCGTCTCGCGCGCGGAGGAGCGAGGCTCCATGAGCGCGCGCAAC 902
 299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
 903 CTGTTCAAGCAATGCTCGCGAGGCGCGCTCGGTGTGATCGCGCGCACGCGCTGGAG 962
 319 GluTyArgLysTyValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyr 338
 963 GAGTACCGAGTACGTGAGAGGAGCGCGCTTCCAGAGGCGGTTCAGCAGGTGTTTC 1022
 339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlyLysTyr 358
 1023 GTCCGCGAGCGCGCTCCGCGCACCATCAGCATCTCGAGGCGCTCAAGGAGAGTAC 1082
 359 GluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378
 1083 GAGGCGCACCAACGCTGTGAGATCCAGGACCGCGCGCTCTGTGCGCGCGCGCTCTCG 1142
 379 AlaArgTyIleThr----- 383
 1143 CGAGGTATACATCAT-GGGTAAGTTGCTCTTCCGCGCTCTTCCCATTTGCCAATGCCAACC 1201
 384 -----GlyArg 385
 1202 CATTTCAAGCTCGAGTTCAATTTGTGCGCATTTCTGCTGTATTTCTTCAGGAGCGC 1261
 386 HisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAlaAsnValArgValGln 405
 1262 CACTGTGCTGCAAGCGATCGACCTCGTGGCGAGCGCGCGCAACGTGAGGTGCGAG 1321
 406 LeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArgMetGlnLeuGluIle 425
 1322 CTGACAGCCAGCAGGAGATCGACAACTCGAGAGGAGAGGATTCAGTGGAGGTG 1381
 426 GluLeuHisAlaLeuGluArgGlyAspLysAlaSerLysAlaArgLeuIleGluVal 445
 1382 GAGCACCGCGCTCGAGAGGAGAGGAGCAAGCAAGCAAGCGCGCGCTTGTGCGAGTTC 1441
 446 ArgLysGluLeuAspLeuArgAspLysLeuGlnProLeuThrMetLysTyArgLys 465

```

1442 AAGAAAGAGCTGGACGATCTGAGGACAAAGCTGCAGCAATTGACGATGAAGTAGTACAGGAAG 1501
466 GluIysGluArgIleAspGluIleArgArgLeuIysGlnIysArgGluGluLeuMetPhe 485
1502 GAGAAAGAGAGGATCGACGAGATCAGGAAGCTGAAGCAGCGCGGAGGAGCTCCAGTTC 1561
486 SerLeuGlnGluAlaGluArgArgTyrAspLeuAlaAlaAspLeuArgTyrGly 505
1562 ACGCTGAGGAGGCGGAGCGCGGATGACCTCGCCCGCTCGCCGACCTCAAGTACGGC 1621
506 AlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAsn 525
1622 CGCTCCAGGAGATCGACGTCGCCATCGCAAGCTCCAG---AGCGAGACGGGAGAAC 1678
526 ValMetLeuThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThr 545
1679 CTGATGCTACCGAGACCGTTCGGCCCGGAGCAGATCCCGAGGTGGTGGCCGCTGAC 1738
546 GlyIleProValThrArgLeuGlyGlnAsnGluIysGluArgLeuIleGlyLeuAlaAsp 565
1739 GGCAATCCCGGTGACCGGCTCGGCCAGAACAGCAGGAGAGGCTCGTCCGCCCTCGCCGAC 1798
566 ArgLeuHisIysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIle 585
1799 AGGTGACACAGAGGGTGTGCGCCAGCGCGGAGCGCTGAGCGCGCTCGCCGAGCGCGTG 1858
586 LeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeu 605
1859 CTGAGTTCGGCGCGGACCTTGGCGCGCGCAGCAGCCACCGCTCGTCTCTCTCTCTCTC 1918
606 GlyProThrGlyValGlyIysThrGluLeuAlaIysAlaLeuAlaGluGlnLeuPheAsp 625
1919 GGTCCCAACCGCGTCCGCAAGACCGAGCTCGCCAAAGCCCTCGCCGAGCAGCTGTTCCGAC 1978
626 AspGluAsnLeuLeuValArgIleAspMetSerGluThrMetGluGlnHisSerValSer 645
1979 GACGAGAACCTGCTCGTCCGCATCGCATGTCGAGTACATGAGCAGCAGCACTCCGTCGCT 2038
646 ArgLeuIleGlyAlaProPro-Gly----- 653
2039 CGCCTCATCGAGACCACTCGGTGGTAGTATCAATCCCAATGTTGTTTTCACAGTTCCTG 2098
654 -----TyrValGly 656
2099 CAACCGAATTTCCGGTGTGACATGGGGAAATTTCTCTCTCTGATGAAGCTACGTCGG 2158
656 YHisGluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLe 676
2159 CCATGAGGAAAGGTGGGAGCTGACGAGCAAGTCCGCGCGGCGGTACAGCGGTGATCCT 2218
676 uPheAspGluValGluIysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAs 696
2219 GTACGACGAGGTGAGAGGCGGACGTCGGGGTGTCAAACCTCTGCTCCAGGTTCCTGA 2278
696 pAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleI 716
2279 CGACGGAGGCTCACCGACGGGCAAGCGCGCACGGTGGACTTCAGGAACACCGGTGATCAT 2338
716 eMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyIysValThrMe 736
2339 CATGACCTCCAACTCCGGCGCGGAGCACTCTCGCCGCGATGGTCCGGCAAGAACTCCAT 2398
736 tGluValAlaArgAspCysValMetArgGlu----- 746
2399 GAAGGTCCCGCGTACCTGTGTCATGAGGAGGTATGATGGCCCTTTGCTTTCACCAACC 2458
747 -----Valar 748
2459 ATGGATTCAATTGCAATGCGATGCTGAATTGATTTTCGATTTTGTGCGAGGTGAG 2518
748 GlyHisPheArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspProLe 768
2519 GAGGCACTTCCGCTCTGAGCTCCTGAAACCGGCTCGACGAGATTGTGATATTCGATCCCT 2578

```

```

QY 768 uSerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLe 788
Db 2579 GTCCACAGCAGCAACTCAGAGAGTCCCGGTCCTCAGATGAAGACGTCGCGCTCCGCT 2638
QY 788 uAlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGl 808
Db 2639 CGCCGCGAGGGGTGTGCGCCTGGCCGCTCACCAGCCGCGCTGGACGTCATTCTCTCTCT 2698
QY 808 uSerTyrAspPro----- 812
Db 2699 GTCCATGATGATCCGCTACCGCAGCAGCATCCAACTATCAGGTGGAAGTTTTTCAAAGAAAT 2758
QY 813 -----ValTyrGlyA 816
Db 2759 CTCGTCGAATTTGATTTCTTAATCTCTCTGTGACTCTGCTACTACGTACAGGTTTATGGG 2818
QY 816 laArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValV 836
Db 2819 CGAGCCGATCAGGAGGTGGATCGAGAAAGAGGTGTCAGCAGCTGTCCAAGATGCTGA 2878
QY 836 alArgGluGluIleAspGluAsnSerThrValTyrIleAspAlaGly-----AlaGlyA 854
Db 2879 TCCAGAGGAGATCGACGAGAACTGCACAGTGTATCATCGCGCGCGCCGCCACAGGACG 2938
QY 854 spLeuValTyrArgValGluSer---GlyGlyLeuValAspAlaSerThrGlyLysLys 873
Db 2939 AGCTGGCTTACAGGTGGACCAACCGCGGGGGTGGTCAACCGCGAGACGGGCGAGAAGT 2998
QY 873 erAspValLeuIleHisIleAlaAsnGlyProLysArg---SerAspAlaAlaGlnAlaV 892
Db 2999 CGGACATCTCATCCAGGTGCCCAACCGCGCGCGCCACCGGAGGAGCAGCGCGCAGCGCG 3058
QY 892 alLysLysMetArgIleGluGluIleGluAspAspAsnGluGlu 907
Db 3059 TGAGAAGATGAGGATCATGGAGACGAGGACGGCATCGACAGAGAG 3105

```

Search completed: February 13, 2004, 04:00:20
Job time : 7680 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

nucleic - nucleic search, using sw model

on: February 12, 2004, 17:31:27 ; Search time 11016 Seconds
(without alignments)
11530.900 Million cell updates/sec

US-09-812-350-30
fect score: 3105
quence: 1 aaagtatcaattttacac.....tctttgggggtttttcata 3105

ring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

2888711 seqs, 20454813386 residues
al number of hits satisfying chosen parameters: 5777422

imum DB seq length: 0
imum DB seq length: 2000000000
t-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rtd.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3105	100.0	3105	8	ATU13949	U13949 Arabidopsis
2	2732.8	88.0	2736	6	AX505540	AX505540 Sequence
3	2732.8	88.0	2736	6	AX589851	AX589851 Sequence
4	2535.8	81.7	6376	8	AF218796	AF218796 Arabidops
5	2532.2	81.6	50821	8	AC020579	AC020579 Arabidops
6	1638.4	52.8	3049	8	SOY8B100HS	L35272 Glycine max
7	1564.4	50.4	3052	8	AF083343	AF083343 Nicotiana
8	1542.4	49.7	1570	8	AY062596	AY062596 Arabidops
9	1345	43.3	3084	8	AF133840	AF133840 Zea mays
10	1334.2	43.0	3058	8	AF174433	AF174433 Triticum
11	1289.2	41.5	3148	8	AF332981	AF332981 Oryza sat
12	1246.8	40.2	3107	8	AF097363	AF097363 Triticum
13	1236.2	39.8	2821	8	AF083344	AF083344 Triticum
14	978.8	31.5	197976	8	ATCHRIV39	AL161539 Arabidops
15	978.8	31.5	206606	8	ATFCAL	Z97336 Arabidops
16	931.4	30.0	2045	8	AF083327	AF083327 Zea mays
17	896.4	28.9	145149	2	AC130611	AC130611 Oryza sat
18	896.4	28.9	162429	2	AC132491	AC132491 Oryza sat
19	888.4	28.6	3109	8	OSA316025	AJ316025 Oryza sat
20	807	26.0	3942	8	AF077337	AF077337 Zea mays
21	750.4	24.2	301150	1	AP003190	AP003190 Clostridi
22	706.4	22.8	10382	1	AE007611	AE007611 Clostridi
23	691.4	22.3	303671	1	AE015937	AE015937 Clostridi
24	679.2	21.9	10700	1	AE010496	AE010496 Fusobacte
25	677.6	21.8	342800	1	AP003598	AP003598 Nostoc sp
26	673.2	21.7	300547	1	AE016954	AE016954 Enterococ
27	668.2	21.5	43980	6	AX413014	AX413014 Sequence
28	668.2	21.5	258650	1	AL596171	AL596171 Listeria
29	668.2	21.5	349980	6	AX417047	AX417047 Sequence
30	666.8	21.5	2776	1	SSU20646	U20646 Synchococc
31	665.6	21.4	290029	1	AE017027	AE017027 Bacillus
32	659.8	21.2	11001	1	AE006383	AE006383 Lactococc
33	656.6	21.1	3247	8	AF449501	AF449501 Paracocci
34	656	21.1	300300	1	AP005373	AP005373 Thermosyn
35	655	21.1	300885	1	AE012559	AE012559 Xylella f
36	653.8	21.1	295050	1	AL591982	AL591982 Listeria
37	653.8	21.1	301442	1	AE016798	AE016798 Vibrio vu
38	653.8	21.1	349980	6	AX641671	AX641671 Sequence
39	652.8	21.0	304708	1	AE017001	AE017001 Bacillus
40	652	21.0	282610	1	RPXX01	AJ235270 Rickettsi
41	651.8	21.0	10092	1	AE003889	AE003889 Xylella f
42	651.6	21.0	2863	1	AF061279	AF061279 Plectonem
43	650.2	20.9	4109	1	AF016634	AF016634 Lactococc
44	649.4	20.9	13313	1	U32767	U32767 Haemophilus
45	649.4	20.9	110000	6	AR274513_08	Continuation (9 of

ALIGNMENTS

RESULT 1
ATU13949
LOCUS Arabidopsis thaliana heat shock protein AtHSP101 mRNA linear PLN 06-JUN-1995
DEFINITION complete cds.
ACCESSION U13949
VERSION U13949.1 GI:537445
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 3105)
AUTHORS Schirmer,E.C., Lindquist,S. and Vierling,E.

TITLE
 JOURNAL
 EDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 TURES
 source

An Arabidopsis heat shock protein complements a thermotolerance defect in Yeast
 Plant Cell 6 (12), 1899-1909 (1994)
 95170291
 7866032
 2 (bases 1 to 3105)
 Direct Submission
 Submitted (24-AUG-1994) Elizabeth Vierling, Biochemistry,
 University of Arizona, Life Sciences South, Tucson, AZ 85721, USA

Location/Qualifiers
 1..3105
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="1-18"
 /tissue_type="leaf"
 /clone_lib="lambda ZAP (Stratagene) cDNA library from heat shock poly(A) RNA; K.W. Helm and E. Vierling, Nuc. Acids Res., 17:7995 (1989)"
 /dev_stage="adult"
 1..3105
 /gene="Athsp101"
 /note="this protein is a member of the highly conserved HSP100/Clp family of proteins, and like other HSP100/Clp proteins, Athsp101 has two consensus ATP-binding domains"
 /codon_start=1
 /product="Athsp101"
 /protein_id="AAA67927.1"
 /db_xref="GI:537446"
 /translation="MNPKFTKNTETATHELAHVAGHAQFPLHLAAGLISDPTG

3' UTR

polyA_site

gene

5' UTR

CDS

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

source

```

1. .2736
/organism="Arabidopsis thaliana"
/db_type="genomic DNA"
/db_xref="taxon:3702"
784 a 481 c 785 g 686 t
GIN
Query Match 88.0%; Score 2732.8; DB 6; Length 2736;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2734; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
163 ATGAATCCAGAGAAATTCACACACAGACAAACGACAGCAATTCGTACAGCTCATGAGCTA 222
1 ATGAATCCAGAGAAATTCACACACAGACAAACGACAGCAATTCGTACAGCTCATGAGCTA 60
223 GCTGTAATGAGGACATGCTCAATTCACCTCTCTGTAATTTAGCTGGTGTCTCTCTCT 282
61 GCTGTGAATGAGGACATGCTCAATTCACCTCTCTCTGTAATTTAGCTGGTGTCTCTCTCT 120
283 GATCCACCGGTATATTTCTCAAGCAATCTCTAGTGGCGGTGGCGAAGCGCAGCTCAA 342
121 GATCCACCGGTATATTTCTCAAGCAATCTCTAGTGGCGGTGGCGAAGCGCAGCTCAA 180
343 TCTGCTGAAAGAGTGATCAATCAAGCCTTGGAAGAGCTTCCTTCAATCTCTCCACCT 402
181 TCTGCTGAAAGAGTGATCAATCAAGCCTTGGAAGAGCTTCCTTCAATCTCTCCACCT 240
403 GATGATATTCAGCGAGTCTAGTCTTATTAAGTCAATTCGTCGCTCAAGCTGCTCAG 462
241 GATGATATTCAGCGAGTCTAGTCTTATTAAGTCAATTCGTCGCTCAAGCTGCTCAG 300
463 AAGTCAAGAGTGATCAATTCGTCGCTGACCAAGTTCATTTAGTTCCTTGAAGAT 522
301 AAGTCAAGAGTGATCAATTCGTCGCTGACCAAGTTCATTTAGTTCCTTGAAGAT 360
523 TCTCAATCAGGATTTGTTGAACGAAGTCGGTGTAGCGACGGCGAGGGTAAAGTCTGAG 582
361 TCTCAATCAGGATTTGTTGAACGAAGTCGGTGTAGCGACGGCGAGGGTAAAGTCTGAG 420
583 GTTGAGAGCTTCGTGGGAAAGAGGAGAAAGTTGAGTTCCTCAGGGGACACAAAT 642
421 GTTGAGAGCTTCGTGGGAAAGAGGAGAAAGTTGAGTTCCTCAGGGGACACAAAT 480
643 TTTCAAGCTTTTAAAGACTTATGGAAGAGATTTGTTGAGCAAGCAGGGAAGCTGATCCT 702
481 TTTCAAGCTTTTAAAGACTTATGGAAGAGATTTGTTGAGCAAGCAGGGAAGCTGATCCT 540
703 GTGATTCGTCGTGATGAGAGATTTAGAGAGTTCGTGAGATTCCTTCAGGAGAACGAAG 762
541 GTGATTCGTCGTGATGAGAGATTTAGAGAGTTCGTGAGATTCCTTCAGGAGAACGAAG 600
763 AACAAATCCTGCTTATTCGAGAGCCAGGAGTTGTTGAAACAGCTGTGTTGAAGTTTA 822
601 AACAAATCCTGCTTATTCGAGAGCCAGGAGTTGTTGAAACAGCTGTGTTGAAGTTTA 660
823 GCACAAAGGATTTGGAAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTCG 882
661 GCACAAAGGATTTGGAAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTCG 720
883 TTGGACATCGGTCGCTAGTTCGTGCTTAATACCGAGGAGAGTTTGAAGAAAGTTG 942
721 TTGGACATCGGTCGCTAGTTCGTGCTTAATACCGAGGAGAGTTTGAAGAAAGTTG 780
943 AAATCTGTTTGAAGAAAGTTGAGGAGCTGAAGCAAGTGAATCTCTTTTATGATGAG 1002
781 AAATCTGTTTGAAGAAAGTTGAGGAGCTGAAGCAAGTGAATCTCTTTTATGATGAG 840
1003 ATTCATTTGGTCTTGGTCTGCGCAAACTGAAGGTCGATGATGAGTCAAGCTAATCTGTC 1062
841 ATTCATTTGGTCTTGGTCTGCGCAAACTGAAGGTCGATGATGAGTCAAGCTAATCTGTC 900
1063 AAGCCATGTTAGCTAGAGGGCAGCTTCATGATTTGGTCTTCAACGCTTGAAGATAC 1122

```

```

Db
901 AAGCCCATGTTAGCTAGAGCGCAGCTTCGATGCAATTTGGTCTCAACGCTTGAAGAATAC 960
Qy
1123 AGGAATATGTTGAGAAAGATGCTGCTTTGAGAGGAGGTTCCAAACAAGCTCTATGTTGG 1182
Db
961 AGGAATATGTTGAGAAAGATGCTGCTTTGAGAGGAGGTTCCAAACAAGCTCTATGTTGG 1020
Qy
1183 GAGCCAAAGTGTGCTGCTGACACCAATAGTATCCTTTAGAGGACTCAAGGAGAGATGAGGGA 1242
Db
1021 GAGCCAAAGTGTGCTGCTGACACCAATAGTATCCTTTAGAGGACTCAAGGAGAGATGAGGGA 1080
Qy
1243 CATCATGTTGTTGGAATCCAAAGACAGAGCTCTTAATAATGCTGCTCAGCTGTCTGCTGT 1302
Db
1081 CATCATGTTGTTGGAATCCAAAGACAGAGCTCTTAATAATGCTGCTCAGCTGTCTGCTGT 1140
Qy
1303 TACATAACTGTTGCTGCGCATTTACCGGATTAAGCAATTTGATTTGTTGATGAGGCTTGTGG 1362
Db
1141 TACATAACTGTTGCTGCGCATTTACCGGATTAAGCAATTTGATTTGTTGATGAGGCTTGTGG 1200
Qy
1363 AATGTGAGAGTCCAGCTTGTATAGTCAACCTGAAGAGATTTGATAACCTTGAAGAGAGAG 1422
Db
1201 AATGTGAGAGTCCAGCTTGTATAGTCAACCTGAAGAGATTTGATAACCTTGAAGAGAGAG 1260
Qy
1423 ATGCAGTGGAAATTTGAACTTCAAGCTTTGGAAGAGAGAGATTAAGCCAGCAAGCT 1482
Db
1261 ATGCAGTGGAAATTTGAACTTCAAGCTTTGGAAGAGAGAGATTAAGCCAGCAAGCT 1320
Qy
1483 CGACTTATAGAGTTCGGAAGAGAGCTTTGATGACCTGAGAGACAGCTTCAGCTCTCAG 1542
Db
1321 CGACTTATAGAGTTCGGAAGAGAGCTTTGATGACCTGAGAGACAGCTTCAGCTCTCAG 1380
Qy
1543 ATGAAATACAGAAAGAGAGAGAGAAATTTGATGAGATTCGAAGCTTTAAACAGAAAGA 1602
Db
1381 ATGAAATACAGAAAGAGAGAGAGAAATTTGATGAGATTCGAAGCTTTAAACAGAAAGA 1440
Qy
1603 GAGAGCTCATGTTTCTTTGAGAGGAGAGAGAGATTAAGCTTCAGAGAGCTGCT 1662
Db
1441 GAGAGCTCATGTTTCTTTGAGAGGAGAGAGAGATTAAGCTTCAGAGAGCTGCT 1500
Qy
1663 GATCTAAGATATTCGCGCAATTCAGAAAGTGGATCTGCAATTTGCCCAAATTTGAAGGAA 1722
Db
1501 GATCTAAGATATTCGCGCAATTCAGAAAGTGGATCTGCAATTTGCCCAAATTTGAAGGAA 1560
Qy
1723 TCTTTGAAAGAGAAATGTTGATGCTCAAGAAAGCTTTGGGCTGGAACACATTTGCTGAGTT 1782
Db
1561 TCTTTGAAAGAGAAATGTTGATGCTCAAGAAAGCTTTGGGCTGGAACACATTTGCTGAGTT 1620
Qy
1783 GTGAGCGTTGAGACAGGATTCAGTTCAGAGACTTTGGCCAAATTCAGAGAGAGAGGTTG 1842
Db
1621 GTGAGCGTTGAGACAGGATTCAGTTCAGAGACTTTGGCCAAATTCAGAGAGAGAGGTTG 1680
Qy
1843 ATTTGCTTCTCTGATAGTTGCAATAAGCGGTTTGTGGGACAGATCAAGCGGTAAATGCA 1902
Db
1681 ATTTGCTTCTCTGATAGTTGCAATAAGCGGTTTGTGGGACAGATCAAGCGGTAAATGCA 1740
Qy
1903 GTTTCTGAGGCAATTCAGAGTCAAGGAGAGAGCTTTGTTAGGCAACAGCAACTGGA 1962
Db
1741 GTTTCTGAGGCAATTCAGAGTCAAGGAGAGAGCTTTGTTAGGCAACAGCAACTGGA 1800
Qy
1963 TCATTTCTTCTTCTGAGCAACTGTTGTTGGCAAACTGAGCTGCCCAAGCTCTTGT 2022
Db
1801 TCATTTCTTCTTCTGAGCAACTGTTGTTGGCAAACTGAGCTGCCCAAGCTCTTGT 1860
Qy
2023 GAGCAGCTGTTGATGATGAACCTCTTAGTTTCGATTTGATGTCGGAATATATGGA 2082
Db
1861 GAGCAGCTGTTGATGATGAACCTCTTAGTTTCGATTTGATGTCGGAATATATGGA 1920
Qy
2083 CAACACTCTCTCTCTGCTCATTGGGACACACAGAGGATGTTGGTCAAGGAGAGT 2142
Db
1921 CAACACTCTCTCTCTGCTCATTGGGACACACAGAGGATGTTGGTCAAGGAGAGT 1980
Qy
2143 GGACAACTAATCTGAGGCTGTGAGGAGCGGACCTTATTTGTTGTCATCTCTTTGATGAGT 2202
Db
1981 GGACAACTAATCTGAGGCTGTGAGGAGCGGACCTTATTTGTTGTCATCTCTTTGATGAGT 2040

```


1183 GAGCAAGTGTGCTGACACCATTTAGTATCTTAGAGAGCTCAGAGGAAGTATGAGGGA 1242
 1021 GAGCAAGTGTGCTGACACCATTTAGTATCTTAGAGAGCTCAGAGGAAGTATGAGGGA 1080
 1243 CATCATGTTGTCGGAATCAAGACAGAGCTCTTATAAATGCTGCTCAGCTCTGCTCGT 1302
 1081 CATCATGTTGTCGGAATCAAGACAGAGCTCTTATAAATGCTGCTCAGCTCTGCTCGT 1140
 1303 TACATPACTGTGTCGCAATTCACGATAAAGCAATGATTTGTTGATGAGGCTTTGCGG 1362
 1141 TACATAAATGTTGTCGCAATTCACGATAAAGCAATGATTTGTTGATGAGGCTTTGCGG 1200
 1363 AATGTGAGAGTCCAGCTTATAGTCACTGGAAGAGATTTGATTAACCTTTGAAGGAGAGG 1422
 1201 AATGTGAGAGTCCAGCTTATAGTCACTGGAAGAGATTTGATTAACCTTTGAAGGAGAGG 1260
 1423 ATGCACTGGAATTAACCTTCAACGCTTTGGAAGGAGAGATTAAGCCAGCAAGCT 1482
 1261 ATGCACTGGAATTAACCTTCAACGCTTTGGAAGGAGAGATTAAGCCAGCAAGCT 1320
 1483 CGACTTATAGAGTGGGAAAGAGCTTGTATGACCTGAGAGCAAGCTTCAAGCTCTCAGG 1542
 1321 CGACTTATAGAGTGGGAAAGAGCTTGTATGACCTGAGAGCAAGCTTCAAGCTCTCAGG 1380
 1543 ATGAATACAGAAAGGAGAAAGAGAGATTTGATGAGATTCGAAGGCTTTAAACAGAAAGA 1602
 1381 ATGAATACAGAAAGGAGAAAGAGAGATTTGATGAGATTCGAAGGCTTTAAACAGAAAGA 1440
 1603 GAAGAGCTCATGTTTTCTTTGAGGAGGAGAGCAAGATATGACCTTTGCAAGAGCTGCT 1662
 1441 GAAGAGCTCATGTTTTCTTTGAGGAGGAGAGCAAGATATGACCTTTGCAAGAGCTGCT 1500
 1663 GATCTAAGATATGCGCAATTCAGAAAGTGGATCTGCAATTTGCCCAACTTGAAGAACT 1722
 1501 GATCTAAGATATGCGCAATTCAGAAAGTGGATCTGCAATTTGCCCAACTTGAAGAACT 1560
 1723 TCTTCTGAAGAGATGTGATGCTCACAGAAACGTTTGGGCTTGAACATTTGCTCAGGTT 1782
 1561 TCTTCTGAAGAGATGTGATGCTCACAGAAACGTTTGGGCTTGAACATTTGCTCAGGTT 1620
 1783 GTGAGCGGTTGACAGGAGATTCAGTACAGAGCTTTGGCCAAATATGAGAGAGAGTTG 1842
 1621 GTGAGCGGTTGACAGGAGATTCAGTACAGAGCTTTGGCCAAATATGAGAGAGAGTTG 1680
 1843 ATTGCTCTGCTGATAGTTCATAGCGGTTTGGGACAGCAATCAGCGTAAATGCA 1902
 1681 ATTGCTCTGCTGATAGTTCATAGCGGTTTGGGACAGCAATCAGCGTAAATGCA 1740
 1903 GTTCTGAGGCAATTCAGGTCAGGCGAGGACTTTGGTAGGCGCACACAGCCAACTGGA 1962
 1741 GTTCTGAGGCAATTCAGGTCAGGCGAGGACTTTGGTAGGCGCACACAGCCAACTGGA 1800
 1963 TCAATCTTATCTTGACCAACTGTGTGTGGCAAACTGAGCTGCGCAAGCTCTTCT 2022
 1801 TCAATCTTATCTTGACCAACTGTGTGTGGCAAACTGAGCTGCGCAAGCTCTTCT 1860
 2023 GAGCAGCTGTTGATCATGAAACCTTTAGTTTGGATTGATATGTCGAATATATGGA 2082
 1861 GAGCAGCTGTTGATCATGAAACCTTTAGTTTGGATTGATATGTCGAATATATGGA 1920
 2083 CAACACTCTGCTCTCGCTCATTTGGGGCACCAAGGATATGTTGGTCAAGGAGAGT 2142
 1921 CAACACTCTGCTCTCGCTCATTTGGGGCACCAAGGATATGTTGGTCAAGGAGAGT 1980
 2143 GGACAACTAAGCTGAGGCTGAGAGGAGGACCTTATGTCATCTCTTATGAGT 2202
 1981 GGACAACTAAGCTGAGGCTGAGAGGAGGAGGACCTTATGTCATCTCTTATGAGT 2040
 2203 GAGAAGCTCATGTTGCTCTTCAACACTCTGCTCAAGTTTGGATGATGCTGATG 2262
 2041 GAGAAGCTCATGTTGCTCTTCAACACTCTGCTCAAGTTTGGATGATGCTGATG 2100

2263 ACAGCGGCAAGCGAGGACAGTCGATTTTCAGAACTCGGTGATATCATCATCAAC 2322
 2101 ACAGCGGCAAGCGAGGACAGTCGATTTTCAGAACTCGGTGATATCATCATCAAC 2160
 2323 CTTGGTCTGAAACACCTCTTTCAGGGCTAACTGGGAAAGTAAACAATGAAAGTGGCCGG 2382
 2161 CTTGGTCTGAAACACCTCTTTCAGGGCTAACTGGGAAAGTAAACAATGAAAGTGGCCGG 2220
 2383 GACTGTGTGATGCGGAGGTGAGGAACACTTCAGACCAAGAGCTCTTGAACAGGCTTGAC 2442
 2221 GACTGTGTGATGCGGAGGTGAGGAACACTTCAGACCAAGAGCTCTTGAACAGGCTTGAC 2280
 2443 GAGATTGTGTGTTCGACCCCTTTTCACATGACCACTTCAGGAAAGTAGCTCGGCTTCAA 2502
 2281 GAGATTGTGTGTTCGACCCCTTTTCACATGACCACTTCAGGAAAGTAGCTCGGCTTCAA 2340
 2503 ATGAAAGACCTTCTGTCGGCTTTCGAAAGAGAGGTTGCTTTGGCAGTCACTGATGCT 2562
 2341 ATGAAAGACCTTCTGTCGGCTTTCGAAAGAGAGGTTGCTTTGGCAGTCACTGATGCT 2400
 2563 GCTTTGGACTATATCTTGGCAGAGGTTTATGACCCGCTGATGCTAGGCTTAAAG 2622
 2401 GCTTTGGACTATATCTTGGCAGAGGTTTATGACCCGCTGATGCTAGGCTTAAAG 2460
 2623 AGATGGATGAGAGAGGTTGTTGACAGAACTCTGCAAGATGTTGCTGAGGAAATC 2682
 2461 AGATGGATGAGAGAGGTTGTTGACAGAACTCTGCAAGATGTTGCTGAGGAAATC 2520
 2683 GATGAAACCTCCTGTTTACATAGATGACGGCTGTTGATCTTGTACCGGTAGAA 2742
 2521 GATGAAACCTCCTGTTTACATAGATGACGGCTGTTGATCTTGTACCGGTAGAA 2580
 2743 AGTGGAGTCTAGTGAAGCTTCAACAGGCAAGCTCAGATGCTGATTCATATGCT 2802
 2581 AGTGGAGTCTAGTGAAGCTTCAACAGGCAAGCTCAGATGCTGATTCATATGCT 2640
 2803 AACGGGCAAGAGAGTGTGATGAGCTCAGGCTGAGGAGAGATGAGGATCGAGGAAATA 2862
 2641 AACGGGCAAGAGAGTGTGATGAGCTCAGGCTGAGGAGAGATGAGGATCGAGGAAATA 2700
 2863 GAAGATGACCATATGAGGAATGATCGAGGATTA 2898
 2701 GAAGATGACCATATGAGGAATGATCGAGGATTA 2736

RESULT 4
 AF218796
 LOCUS Arabidopsis thaliana heat shock protein 101 (HSP101) gene, complete cds.
 DEFINITION Arabidopsis thaliana 6376 bp DNA linear PLN 13-APR-2000
 ACCESSION AF218796
 VERSION AF218796.1 GI:6715467
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
 1 (bases 1 to 6376)
 Hong, S.W. and Vierling, E.
 Mutants of Arabidopsis thaliana defective in the acquisition of
 tolerance to high temperature stress
 Proc. Natl. Acad. Sci. U.S.A. 97 (8), 4392-4397 (2000)
 20226114
 PUBMED 10760305
 2 (bases 1 to 6376)
 Hong, S.-W. and Vierling, E.
 Direct Submission
 Submitted (22-DEC-1999) Biochemistry, University of Arizona, 1007
 E. Lowell St., Tucson, AZ 85721, USA
 Location/Qualifiers
 1..6376
 /organism="Arabidopsis thaliana"

```

/molecule="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="ngalli-ATATPASE"
<1132..>4279
/gene="HSP101"
/join(<1132..2281,2366..2547,2688..3313,3398..3875,
3980..>4279)
/gene="HSP101"
/product="heat shock protein 101"
/join(1132..2281,2366..2547,2688..3313,3398..3875,
3980..>4279)
/gene="HSP101"
/feature="clpB"
/codon_start=1
/product="heat shock protein 101"
/protein_id="AAF26423.1"
/db_xref="GI:6715468"
/translation="MNPBKFTHKNTIATHELAHVAGHAQFTPLHLAGALISDPTG
IPQAISSAGGENAQAQSRVINOALKKLPQSPPPDDIPASSSLIKVIRRAQAOKS
RGDTHLVDQLIMGLLEDSQJDLANEVGVATARVSEKLRKEGKGVESAGSDYN
FOALTYGRDLVQAGKLDPIVGRDEIRRVIRLSRTKNPVLIGPVGKTAIVE
GLAQRIVGDVNSUTDVLISLMDGALVAGAKYGEPEERLKYKVEVEDAEKRVIL
FIDEHLVIGAGKTGSDAANLFPMLARQLRCIGATTLERYKYVEKDAAFERF
QQVYVAEPSPVDITISILRLAKEYEGHGVRIQDRLINAAQLSARYITGRHLPKAI
DLVDEACANVRVOLDSPREIDNLRKMOLEIEHLALEREKQASKARLLEVRKELD
DLRDLKLPMTKYRKEKRIEIRLKKRBEMLFSLQEAERVDLARAADLRVGAIQ
EVSALIQLEQSSSENWMLTENUGPEHIAEVSVRWGTGIPVTRLOONKRLIGLADR
LHKRVQGNQANVAESALRSRAGLRPQPTGSFLGPTGVGKTELAKALAEQLF
DENLILVRIDSMYEQHSVRLIGAPPGYHEEGQLEAVRRPVCVILFDEVERK
AHVAVENTLLQVLDGRITDGGRTVDFPNSVIIMTSLNGLAHLGLTGRVTVMEVAR
DCVMEVRKRPRLDELTVDPDPLSHDQLKVARLQMDKVAIRLAEGRVALAVT
DAALDYILAESVDYVYKGLRIRRMKKVTELKSMVVEEIDENSFTYIDAGADLV
YRVESGLVDASTGKSDVLIHANGPKRSDRAQVKKMRIBIEDDDNEEMIED"
1914 a 1150 c 1470 g 1842 t

```

```

377 AGCTTCCTTCACAATCTCCCTCCACCTGATGATATTCAGCGAGTTCCTAGTCTTATTAAGG 436
1346 AGCTTCCTTCACAATCTCCCTCCACCTGATGATATTCAGCGAGTTCCTAGTCTTATTAAGG 1405
437 TCATTCTCTGCTGCTCAAGCTGCTCAGAAAGTCAAGAGTATATCTCAATTTGGCTGTTGAAC 496
1406 TCATTCTCTGCTGCTCAAGCTGCTCAGAAAGTCAAGAGTATATCTCAATTTGGCTGTTGAAC 1465
497 AGTTGATTATGGTCTCTCTTGAAGATTCCTCAATCAGGAGTATTTGTTGAACGAAGTGGTG 556
1466 AGTTGATTATGGTCTCTCTTGAAGATTCCTCAATCAGGAGTATTTGTTGAACGAAGTGGTG 1525
557 TAGCGACGGCAGGGTAAAGTCTGAGGTTGAGAAAGTTCCTGCGGAAGAGGAGGAAGAAAG 616
1526 TAGCGACGGCAGGGTAAAGTCTGAGGTTGAGAAAGTTCCTGCGGAAGAGGAGGAAGAAAG 1585
617 TTGAGAGTGTCTTCAGGGGACACAAATTTTCAAGCTTTAAAGATTTATGGAAGAGATTTGG 676
1586 TTGAGAGTGTCTTCAGGGGACACAAATTTTCAAGCTTTAAAGATTTATGGAAGAGATTTGG 1645
577 TTGAGCAAGCAGGGAAGCTTGATCTGCTGATTTGCTGCTGATGAGGAGATTTAGAGAGTGG 736
1646 TTGAGCAAGCAGGGAAGCTTGATCTGCTGATTTGCTGCTGATGAGGAGATTTAGAGAGTGG 1705
737 TGAGGATTTCTTTCAGGAGAAACGAAGAACTCTGCTGCTGATTTGAGAGAGCAGGAGTGG 796
1706 TGAGGATTTCTTTCAGGAGAAACGAAGAACTCTGCTGCTGATTTGAGAGAGCAGGAGTGG 1765
797 GTAAACAGCTGTGGTTGAAGTTTATGACAAAGAGTGTGGAAGAGATTTGCCCAACA 856
1766 GTAAACAGCTGTGGTTGAAGTTTATGACAAAGAGTGTGGAAGAGATTTGCCCAACA 1825
857 GTCTTACTGATGTGAGATTAATTTCTGTTGAGACATGGTGGTGGTGGTGGTGGTGGTGGT 916
1826 GTCTTACTGATGTGAGATTAATTTCTGTTGAGACATGGTGGTGGTGGTGGTGGTGGTGGT 1885
917 ACCGAGGAGATTTGAAGAAAGGTTGAAATCTCTTTTGAAGAAAGTGTGAGGAGCCTGAAG 976
1886 ACCGAGGAGATTTGAAGAAAGGTTGAAATCTCTTTTGAAGAAAGTGTGAGGAGCCTGAAG 1945
977 GCAAAGTGAATCTCTTTATTTATGATGAGATTCATTTGTTGCTGCTGCTGCTGCTGCTGCTG 1036
1946 GCAAAGTGAATCTCTTTATTTATGATGAGATTCATTTGTTGCTGCTGCTGCTGCTGCTGCTG 2005
1037 GGTGATGATGAGTGAATCTCTTTCAAGCCCATGTTAGCTAGAGGAGCAGCTTCGATGCA 1096
2006 GGTGATGATGAGTGAATCTCTTTCAAGCCCATGTTAGCTAGAGGAGCAGCTTCGATGCA 2065
1097 TTGCTGCTACACGCTTGAAGAAATACAGGAATATGTTGAGAAAGATGCTGCTGCTGCTGCTG 1156
2066 TTGCTGCTACACGCTTGAAGAAATACAGGAATATGTTGAGAAAGATGCTGCTGCTGCTGCTG 1215
1157 GGAAGTTCCAAAGTCTATGTTGCGGAGCAGCAAGTGTGCTGACACATAGTATCTCTTA 1216
2126 GGAAGTTCCAAAGTCTATGTTGCGGAGCAGCAAGTGTGCTGACACATAGTATCTCTTA 2185
1217 GGAAGTTCCAAAGTCTATGTTGCGGAGCAGCAAGTGTGCTGACACATAGTATCTCTTA 1276
2186 GGAAGTTCCAAAGTCTATGTTGCGGAGCAGCAAGTGTGCTGACACATAGTATCTCTTA 2245
1277 TAAATGCTGCTCAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1311
2246 TAAATGCTGCTCAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2305
1312 TAAATGCTGCTCAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2365
2306 GCTGATGCTTTATGCTTATATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2365
1313 GTGGGCAATTTACCGGATTAAGCAATTTGATTTGGTTGATGAGGCTTGTGCGATGTGAGAG 1372
2366 GTGGGCAATTTACCGGATTAAGCAATTTGATTTGGTTGATGAGGCTTGTGCGATGTGAGAG 2425

```

Query Match

81.7%; Score 2535.8; DB 8; Length 6376;

Best Local Similarity 85.7%; Pred No. 0;

Matches 3103; Conservative 0; Mismatches 2; Indels 516; Gaps 5;

```

1 AAAGTATCAATTTTACAACTACCGCTATATCTGCTGATTCCTCTGCAAAAGAGAA 60
866 AAAGTATCAATTTTACAACTACCGCTATATCTGCTGATTCCTCTGCAAAAGAGAA 925
61 GACTTTTACGAGAGAGAGTCTCTGCTGATTCGAGAACTCAAGAAACCAACCCAG 120
926 GACTTTTACGAGAGAGAGTCTCTGCTGATTCGAGAACTCAAGAAACCAACCCAG 985
121 TTCTCATATATCGTTTAA----- 138
986 TTCTCATATATCGTTTAAAGTAAATGATCGGCAACATCTGTTCTCATTTGTGTGTTTT 1045
139 ----- 138
1046 TGTGTTTGTGTTAGGGTTTACAAAGATACCTGAGATTAGTTTTTTTTTTTTTTTTTTT 1105
139 --AGGATTACAAAGCTAATCGAGATGAATCCAGAGAAATTCACACACAGCAACAG 196
1106 TCAGGATTTACAAAGCTAATCGAGATGAATCCAGAGAAATTCACACACAGCAACAG 1165
197 AGCAATTTCTACAGCTCATGAGCTAGCTGTGAATGCGAGGACATGCTCAATTCCTCTT 256
1166 AGCAATTTCTACAGCTCATGAGCTAGCTGTGAATGCGAGGACATGCTCAATTCCTCTT 1225
257 TGCATTTAGCTGGTGTGATCTCTGATCCACCGGTATATTTCTCTCAAGCAATCTCTA 316
1226 TGCATTTAGCTGGTGTGATCTCTGATCCACCGGTATATTTCTCTCAAGCAATCTCTA 1285
317 GTCCGGTGGCAGAGACGAGCTCAATCTGCTGAAAGAGTGTGATCAATCAAGCTTGAAGA 376
1286 GTCCGGTGGCAGAGACGAGCTCAATCTGCTGAAAGAGTGTGATCAATCAAGCTTGAAGA 1345

```

E COUNT
GIN

3506	Db		CAC	TCTGCTC	CAAG	TTTTCGG	ATG	TGG	ATG	CAG	CGGG	CAAG	CGCAGG	CAG	CTCGA	3565																	
2289	Qy		TTT	CAGGAA	CTCGGT	GAT	AAT	CAT	GAC	AT	CAAA	CC	TGGT	GCT	GAA	CA	CACT	CTCT	TGCAGG	2348													
3566	Db		TTT	CAGGAA	CTCGGT	GAT	AAT	CAT	GAC	AT	CAAA	CC	TGGT	GCT	GAA	CA	CACT	CTCT	TGCAGG	3625													
2349	Qy		GCT	AACT	CGGAAA	AGT	AA	CAAT	AGG	AGT	GCC	CGG	GA	CTG	CT	GAT	AT	CGG	GAG	GTGAGAA	2408												
3626	Db		GCT	AACT	CGGAAA	AGT	AA	CAAT	AGG	AGT	GCC	CGG	GA	CTG	CT	GAT	AT	CGG	GAG	GTGAGAA	3685												
2409	Qy		AC	CTT	CAG	AC	CAG	AG	CT	CTT	GAA	CAG	GG	CTT	GAC	GAG	AT	CTG	TG	TTT	CGAC	CC	CC	TTT	C	2468							
3686	Db		AC	CTT	CAG	AC	CAG	AG	CT	CTT	GAA	CAG	GG	CTT	GAC	GAG	AT	CTG	TG	TTT	CGAC	CC	CC	TTT	C	3745							
2469	Qy		AC	AT	GAC	CA	CGT	TTG	CAG	GA	AA	GT	AG	CT	CGG	CTT	CA	AA	T	GA	AA	GA	C	GT	TG	CT	GT	CGG	CTT	GC	2528		
3746	Db		AC	AT	GAC	CA	CGT	TTG	CAG	GA	AA	GT	AG	CT	CGG	CTT	CA	AA	T	GA	AA	GA	C	GT	TG	CT	GT	CGG	CTT	GC	3805		
2529	Qy		TG	AA	GAG	GAG	AT	G	CTT	TCG	CA	GT	AG	CT	CA	CT	GT	CTG	CT	TTG	G	AC	T	A	T	A	T	CTT	TG	G	CAG	AG	2588
3806	Db		TG	AA	GAG	GAG	AT	G	CTT	TCG	CA	GT	AG	CT	CA	CT	GT	CTG	CT	TTG	G	AC	T	A	T	A	T	CTT	TG	G	CAG	AG	3865
2589	Qy		TT	AT	GA	CCCC	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2597	
3866	Db		TT	AT	GA	CCCCGGT	AA	GT	CT	CTT	CAC	AT	GAT	GC	AG	AT	G	CAT	A	T	A	T	T	G	T	G	A	T	CT	CT	CAG	3925	
2598	Qy		-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2604	
3926	Db		CT	GG	TTT	GT	CT	GT	CA	CT	CA	CT	TTG	GG	TT	GT	CTT	CTT	TT	TTG	TT	GG	TT	GG	TT	GG	TT	GG	TT	GG	TT	3985	
2605	Qy		GG	CT	TA	GG	CT	TA	AA	GG	AG	AT	GG	AT	GG	AT	GG	AT	GG	AT	GG	AT	GG	AT	GG	AT	GG	AT	GG	AT	GG	AT	2664
3986	Db		GG	CT	TA	GG	CT	TA	AA	GG	AG	AT	GG	AT	GG	AT	GG	AT	GG	AT	GG	AT	GG	AT	GG	AT	GG	AT	GG	AT	GG	AT	4045
2665	Qy		GT	TG	CGT	GAG	GA	AT	CG	AT	GA	AA	CT	CC	AT	CT	TTT	TA	CA	TAG	AT	G	AT	G	AT	G	AT	G	AT	G	AT	G	2724
4046	Db		GT	TG	CGT	GAG	GA	AT	CG	AT	GA	AA	CT	CC	AT	CT	TTT	TA	CA	TAG	AT	G	AT	G	AT	G	AT	G	AT	G	AT	G	4105
2725	Qy		CTT	CT	GT	AC	CGG	TAG	AA	GT	GG	AG	CT	CT	AG	TG	AG	CG	CT	CA	A	CAG	CA	G	A	G	A	G	A	G	A	G	2784
4106	Db		CTT	CT	GT	AC	CGG	TAG	AA	GT	GG	AG	CT	CT	AG	TG	AG	CG	CT	CA	A	CAG	CA	G	A	G	A	G	A	G	A	G	4165
2785	Qy		GT	CT	GA	TT	CAT																										

AC020579 50821 bp DNA linear PLN 19-JAN-2001
Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence,
complete sequence.

Qy	2120	-----	2119
Db	15285	TATCTTTTAGTTTTTCAATATCTTCTGAATTATGAAATAATTTTTTCTTACTTACAGATTA	15344
Qy	2120	-----GGTATGCTTGGTCACGAGGAAGGTGGACAACTAACTCAGGCGTGTGAGG	2166
Db	15345	CTACACTTTCTCAGGTATGTTGGTCAAGAGGAAGGTGGACAACTAACTCAGGCGTGTGAGG	15404
Qy	2167	AGCGGACCTTAATTGTGTCTATATCTTTTGTATGAAGTGGAGAAAGGCTCATGTTGCTGTCTTC	2226
Db	15405	AGCGGACCTTAATTGTGTCTATATCTTTTGTATGAAGTGGAGAAAGGCTCATGTTGCTGTCTTC	15466
Qy	2227	AACACTCTGCTCCAGATTTTGGATGATGTCGATTGACAGACGGGCAAGGCAAGGACAGATC	2286
Db	15465	AACACTCTGCTCCAGATTTTGGATGATGTCGATTGACAGACGGGCAAGGCAAGGACAGATC	15524
Qy	2287	GATTTTCAGGAACCTCGGTGATAATCATGACATCAAACTTTGGTGTCTGAACTCTCTTGCA	2346
Db	15525	GATTTTCAGGAACCTCGGTGATAATCATGACATCAAACTTTGGTGTCTGAACTCTCTTGCA	15584
Qy	2347	GGGCTTAACCTGGGNAAGTAACAATGGAAGTGGCCCGGACTGTGTGATGCGGAGGTGAGG	2406
Db	15585	GGGCTTAACCTGGGNAAGTAACAATGGAAGTGGCCCGGACTGTGTGATGCGGAGGTGAGG	15644
Qy	2407	AAACACTTTCAGACACAGAGCTCTTTGAACAGGCTTGACGAGATTTGGTGTTCGACCCCTT	2466
Db	15645	AAACACTTTCAGACACAGAGCTCTTTGAACAGGCTTGACGAGATTTGGTGTTCGACCCCTT	15704
Qy	2467	TGACATGACAGTTTGAGGAAAGTAGCTCGGCTTCAAAATGAAGAGCTTGCTGTCTCGGCTT	2526
Db	15705	TGACATGACAGTTTGAGGAAAGTAGCTCGGCTTCAAAATGAAGAGCTTGCTGTCTCGGCTT	15764
Qy	2527	GCTGAAAGAGGAGTTGCTTTTGGCAGTCACTGATGCTGCTTTTCGACTATATCTTTGCACAG	2586
Db	15765	GCTGAAAGAGGAGTTGCTTTTGGCAGTCACTGATGCTGCTTTTCGACTATATCTTTGCACAG	15824
Qy	2587	AGTTATGACCC-----	2597
Db	15825	AGTTATGACCCGGTAAGTCTCTTTCACATGATGCAGAAATGCATATCATTTGATCTCTCA	15884
Qy	2598	-----GGTGT 2602	
Db	15885	GGCTGTTTGTCTGTGTGACTGACTTGTGTGTGCTTCTTTGTTTGTGTGTTAGGTGT	15944
Qy	2603	ATGGTCTAGGCGCTATAAGGAGATGGATGGAGAGAAAGGTGTGTGACAGAACTGTCAAAGA	2662
Db	15945	ATGGTCTAGGCGCTATAAGGAGATGGATGGAGAGAAAGGTGTGTGACAGAACTGTCAAAGA	16004
Qy	2663	TGTTTGTGTGGTAGGAAATTCGATGAAAACTCCACTGTTTACATAGATGACGGCGTGTGTG	2722
Db	16005	TGTTTGTGTGGTAGGAAATTCGATGAAAACTCCACTGTTTACATAGATGACGGCGTGTGTG	16064
Qy	2723	ATCTCTGTGTACCGGGTAGAAAAGTGGAGGTCTAGTGGACGCTTCAACAGCGAAGAGTCAAG	2782
Db	16065	ATCTCTGTGTACCGGGTAGAAAAGTGGAGGTCTAGTGGACGCTTCAACAGCGAAGAGTCAAG	16124
Qy	2783	ATGTGCTGATTCATATTCGTAAACGGGCCAAAGAGAAGTGAATCGACTCAGGCGGTGAAGA	2842
Db	16125	ATGTGCTGATTCATATTCGTAAACGGGCCAAAGAGAAGTGAATCGACTCAGGCGGTGAAGA	16184
Qy	2843	AGATGAGGATTCGAGGAAATPAGAAGATGACGATTAATGAGGAAATGATCGAGGATTAAGTTC	2902
Db	16185	AGATGAGGATTCGAGGAAATPAGAAGATGACGATTAATGAGGAAATGATCGAGGATTAAGTTC	16244
Qy	2903	TTGACCTCCAAAGTGTGAGTCTCTTTTGAGGTGTGTGATTTGATGACTTTCAGCTTTTCATG	2962
Db	16245	TTGACCTCCAAAGTGTGAGTCTCTTTTGAGGTGTGTGATTTGATGACTTTCAGCTTTTCATG	16304
Qy	2963	CTCTGTTTTGGTTTTGACTCTTTTGGTAGACTATATGATGGTGGAACTGTGTAAGAAAGTCT	3022
Db	16305	CTCTGTTTTGGTTTTGACTCTTTTGGTAGACTATATGATGGTGGAACTGTGTAAGAAAGTCT	16364

1358 GTGCAATGTGAGAGTCCAGCTTGATAGTCAACCTGAAGAGATTGATAA CTTGAAAGGA 1417
1334 GTGCAATGTTAGGTTCAACTTGATAGTCAACCTGAGGAAATTGATAA CTTGAAAGGA 1393
1418 AGAGATGACAGCTGAAATTAACCTTCAGCTTGGAAAGGAGAGATTAAGCCAGCA 1477
1394 AGAAATGCACTAGAGTGAAGTGAATCTCATGCTTGGAGAAAGAGAAAGCAAGCTAGCA 1453
1478 AAGCTGCACTTATAGAGTGGGAAAGAGCTTGATGACCTGAGAGACAAGCTTCAGCCTC 1537
1454 AAGCCCGTCTTGTTGAGTGGGAAAGAGCTTGATGACCTTGGAGGACAGCTTCAGCCTT 1513
1538 TCAGATGAATAACAGAAAGAGAGAGAGAAATTTGATGAGATTCGAAAGCTTTAAACAGA 1597
1514 TGATGATGAAGTACCGAAAGAGAGAGAGAGGTTGATGAGATTCGAAAGCTCAAGAAGA 1573
1598 AAGAGAGAGCTCATGCTTTCTTTGAGGAGGAGAGAGAGAGAGATGACCTTGGAGAG 1657
1574 AAGAGAGAGCTTCCTTTTCTTTGAGGAGGAGAGAGAGAGAGATGATCTGGCTAGAG 1633
1658 CTGCTGATTAAGATATGGGCAATTCAGAAAGTGGAACTCTGCAATTTGCCCACTTGAAG 1717
1634 CTGCAAGCTCGGATATGGAGCAATTCAGAGGTGGAATCTGCAATACAACTTGAAG 1693
1718 GAATCTCTCTGAAGAGAAATGTGATGCTCAAGAAACGTTGGGCTTGAACACATTTGCTG 1777
1694 GGA---GCACTGAAGAGAACTCTGATTTGACTGAATCTGTTGACCGGACCAATAGCTG 1750
1778 AGGTTGAGCCCTTGACAGGAGATTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1837
1751 AGGTTGAGCCCTTGACAGGAGATTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1810
1838 GGTGATTTGCTTTGCTGATAGTGTGATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1897
1811 GATTGATTTGAGCTTTGATGAGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1870
1898 ATGAGTTCTGAGGCAATTCAGGTTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1957
1871 ATGCTGTTGCTGAGGCTGTTGATGAGATCTAGAGCTGAGAGCTGAGGAGAGAGAGAG 1930
1958 CTGATCATTTCTTTCTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2017
1931 CTGCTCATTTCTTTCTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1990
2018 TTGCTGAGGAGCTTTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2077
1991 TTGCTGAGGAGCTTTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2050
2078 TGGAAACAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2137
2051 TGGAGGAGGAGCTCGGTTTCAAGATTTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2110
2138 AAGGTGAGCAACTAACTGAGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2197
2111 AAGGTGAGCAACTAACTGAGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2170
2198 AAGGTGAGGAGGCTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2257
2171 AAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2230
2258 GATTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2317
2231 GGTAACTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2290
2318 CAACCTTGTGCTGAGCACTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2377
2291 CAATCTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2350
2378 CCGGAGCTGTGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2437
2351 CTGGGATCGGTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2410

QY 2438 TTGACGAGATTTGCTGTTGCTGACCCCTTTTACATGACCCAGTTGAGGAAAGTAGTCCGC 2497
Db 2411 TTGATGAATTTGCTGTTGCTGACCCCTTTTACATGACCCAGTTGAGGAAAGTTGCAAGC 2470
QY 2498 TTCAATGAAGAACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2557
Db 2471 TACAATGAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2530
QY 2558 ATGCTGCTTTGAGCTATATCTTTGAGGAGGAGGATGATGACCCGCTGATGCTGCTGCTG 2617
Db 2531 ATGAGCACTAGACTATATCTTTTCCGAGAGCTATGATCCGCTGATGCTGCTGCTGCTG 2590
QY 2618 TAAGGAGATGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2677
Db 2591 TAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2650
QY 2678 AATCGATGAAGAACTTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2731
Db 2651 AGATTTGATGAGAAATTCACCGGTTTACATAGATGCTGAGACCCCAATGGGCGGAGTTGGTCT 2710
QY 2732 ACCGGGTAGAA---AGTGGAGGCTAGTGGAGGCTTCAACAGGCAAGAGTCAAGATGTC 2788
Db 2711 ACCGCTGAGAAAGAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2770
QY 2789 TGATTCATATTTGCTTAAACGGGC---CAAAGAGAGAGTGTGCTGAGCTCAGGCGGTGAAGAAGA 2845
Db 2771 TGATTCAGATACCTTAATGAGCATGACACCTTAACTGATGCTGCTGCTGCTGCTGCTG 2830
QY 2846 TGAGATCGAGGAAATGAGAGATGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2903
Db 2831 TGAGATTTGAGGAAATTTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2888

RESULT 7
AF083343
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
FEATURES
source
gene
CDS

AF083343
Nicotiana tabacum heat shock protein 101 kDa (HSP101) mRNA,
complete cds.
AF083343
AF083343.2 GI:11561805
Nicotiana tabacum (common tobacco)
Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 3052)
Wells,D.R., Tanguay,R.L., Le,H. and Gallie,D.R.
HSP101 functions as a specific translational regulatory protein
whose activity is regulated by nutrient status
Genes Dev. 12 (20), 3236-3251 (1998)
99033076
9784498
2 (bases 1 to 3052)
Gallie,D.R., Tanguay,R.L. and Wells,D.R.
Direct Submission
Submitted (10-AUG-1998) Biochemistry, University of
California-Riverside, Riverside, CA 92521-0129, USA
3 (bases 1 to 3052)
Gallie,D.R., Tanguay,R.L. and Wells,D.R.
Direct Submission
Submitted (06-DEC-2000) Biochemistry, University of
California-Riverside, Riverside, CA 92521-0129, USA
Sequence update by submitter
On Dec 6, 2000 this sequence version replaced gi:3982830.
Location/Qualifiers
1..3052
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/db_xref="taxon:4097"
1..3052
/gene="HSP101"
101..2830

QY 1 AAAGTTATCAATTATTCACACATTTACCGCTATAATCTGCTTGATTCTCTGCCAAAAGAGAA 60
|||||
Db 27 AAAGTTATCAATTATTCACACATTTACCGCTATAATCTGCTTGATTCTCTGCCAAAAGAGAA 86

QY 61 GACTTTTTACCGAAGAAGACTCCTCTGGCTCATTTGAAGAAACTCAACGAAACAACACCAG 120

AY062596 1570 bp mRNA linear PLN 25-NOV-2001
Arabidopsis thaliana heat shock protein 101 (At1g74310; F1017.2)
mRNA, partial cds.
AY062596

ILRQAITGASGGDGAAGDSFERVLNNSLKKLPSQSPPPDSVPASTALI KVI RRAQSAQ

Query Match 43.3%; Score 1345; DB 8; Length 3084;
Best Local Similarity 69.1%; Pred. No. 0;
Matches 1904; Conservative 0; Mismatches 835; Indels 18; Gaps 4;
GIN
SE COUNT 670 a 874 c 1057 g 483 t

141 GATTACAAAGCTAATCAAGATGAATCAAGAGAAATTCACACACAAAGACAAACGAGAC 200
123 GAGTAGAAG 182
201 AATTGCTACAGCTCAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 260
183 GATCGTGGGGGCGACAGAGATGCGGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 242
261 TTTAGCTGCTGCTTGTGATCTGATCCACCGGTATATTTCTCAGCAATCTCTAGTGC 320
243 CTTGGCGGCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302
321 CGGTGCGAGAGAGCGAGC-----TCAATCTGCTGAAAGAGTGAATCAATCAAGCTTGAA 374
303 GTGCGGGGCGAGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 362
375 GAAGCTTCTTCAATCTCTCCACCTGATGATATTCAGGAGGTTCTAGTCTTATTA 434
363 GAAGCTGCGGTGCGAGTCTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 422
435 GTCATCTGCTGCTCAGCTGCTCAGAGTCAAGAGTCAAGAGTCAATCTGCTGCTGCTG 494
423 GGTCACTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 482
495 CCAGTTGATTATGGGTCTCTTGAAGATCTCAATCAAGGATTTGTTGAAGAGAGTGG 554
483 CCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 542
555 TGTAGCGAGCGGCGGAGGTAAGTCTGAGGTTGAGAGTCTGAGAGTCTGAGAGTCTGAG 614
543 CGTGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 602
615 AGTTGAGAGTCTTCAAGGAGGAGACAAATTTTCAAGCTTTTAAAGACTTATGGAAGAT 674
603 CGTGGAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 662
675 GGTGAGCAGAGCAGGAGGAGTCTGATCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 734
663 CGTTGAGCAGGCGGCGGAGCTGAGACCCGCTCATCGGCGGAGGAGATCCGCGGCT 722
735 CGTGAGGATCTTTCAG 794
723 GGTGCGCATCTCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 782
795 TGTGAG 854
783 TGGCAG 842
855 CAGTCTTACTGATGAGAGATTAATTTCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 914
843 TAACCTCTCGAGTCCGCTCATCGGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902
915 ATACGAG 974

903 GTACCGGCGGCGAGTTCCAGGAGCGGCTCAAGCGCGTCTCAAGGAGGTGGAAGAGCGCGA 962
975 AGGCAAGTGTATCTCTTTATGATGAGATTCATTTGGTCTTGGTCTGGCAAACTGA 1034
963 GGGGAAGGTCTATCTCTTCTATCGAGAGATACACCTGCTCTGGCGCGGCGAGAGCGGA 1022
1035 AGGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1094
1023 GGGTTCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1082
1095 CATTTGGTCTAAGCGTCTGAAGATAAGAGATAAGAGATAAGAGATAAGAGATAAGAG 1154
1083 CATCGCGCGCACCGCTGAGGAGTACCGCAAGTACGTGAGAGAGAGAGAGAGAGAGAG 1142
1155 GAGGAGGTTCCAAAGCTCTATGTTGGGAGCGGAGTGTGCTGAGACCAATGATATCTCT 1214
1143 GCGCGGTTCCAGCAGGTTGTTGTCGCGGAGCGGAGCGTCCCGGACACCGCTCAGCATCT 1202
1215 TAGAGGACTCAAGGAGAGATGAGGAGCATCATGTTGTCGGAATCCAAAGACAGAGCTCT 1274
1203 GAGGAGACTCAAGGAGAGATGAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1262
1275 TATAATGCTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1334
1263 CGTGGTCCGCGCACAGCTATCCGCGAGGTACATCATGGGTGCGGCGGCGGCGGCGGCG 1322
1335 AATTGATTGGTGTGATGAGGCTGTCGGAATGTCGAGTCCAGCTTGTATGATCAACTGA 1394
1323 CATAGACCTGGTGAAGAGGCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1382
1395 AGAGATTGATAAACCCTTCAAGGAGAGAGATGAGGAGTCCAGCTTGGAAATTTGAACTT 1454
1383 GGAGATTGATAAACCCTTGAAGAGAGAGAGATCCAGCTTGTAGTTGAGCTTCAAGCGCT 1442
1455 AAGGAGAGAGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1514
1443 GAAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1502
1515 CTTGAGAGAGAGAGGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1574
1503 TCTGAGGAGAGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1562
1575 TGAGATTGAGAGGCTTAAACAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1634
1563 TGAGATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1622
1635 ACAGAGATATGACCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1694
1623 GCGCGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1682
1695 ATCTGCAATTTGCCCAACTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1754
1683 GCGCGGATCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1739
1755 CGTTGGGCTTGAACACATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1814
1740 GCTGCGGCTTGAACAAATTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1799
1815 ACTTGGGCGAAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1874
1800 GCTTGGGCGAAGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1859
1875 TGTGGGAGAGAGATCAAGCGGTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1934
1860 GGTGCGGCGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1919
1935 ACTTGGTAGGCGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1994
1920 TCTTGGCAGGCGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1979
1995 CAAAGCTGAGGCTCGCGAAGGCTCTTGTGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2054

1980 GAAACGAGCTAGCCAGCCCTAGCCGACAGCTGTTCGACGAGCAACCTCTTGT 2039
 2055 TCGGATGATATGTCGGAATATATGAAACAACTCTGTCTCTCGCTCATTTGGGCGACC 2114
 2040 CCGCATCGACATGTCGAGTACATGAGCAGCACTCGTTGCTCGCTCATCGGACACC 2099
 2115 ACCAGGATATGTTGTCACGAGGAAAGTGAACAACTGAGGCTGTGAGAGGCGACC 2174
 2100 GCCTGGCTAGCTCGGCATCAAGAGGCTGGGAGCTGACTGACAAAGTGAAGGAGGCG 2159
 2175 TTATGTCATCTCTTTCATGAGTGAAGGAGGCTCATGTTGCTGCTTCAACACTCT 2234
 2160 GTACAGGCTGATCTGTTTCAGAGGTCGAGAGGCGCCATGTCCGCTGTCAACACCT 2219
 2235 GCTCCAGTTTTCGATGATGTCATGACGAGGCGGAGGAGGAGGAGGAGGAGGAGG 2294
 2220 GCTCCAGTCTTCGACGAGGAGGTTGAGGATGGGAGGAGGAGGAGGAGGAGGAGG 2279
 2295 GAATCGGTGATATCATGACATCAACCTTGTGTCGACACCTCTCTGAGGAGGAGG 2354
 2280 GAACACCGTGTATCATGACATCAACCTTGTGTCGACACCTCTCTGAGGAGGAGG 2339
 2355 TGGGAGGATGAACATGAGTGGCGGCGGAGTGTGATGCGGAGGAGGAGGAGGAGG 2414
 2340 GGGCAAGACTCCATGAGGTCGCTCGGATCTGGTATGCGAGGAGGAGGAGGAGG 2399
 2415 CAGACGAGGCTCTTGAACAGGCTTGCAGGAGTGTGTTGTCGACCCCTTTTCATCA 2474
 2400 CCGACCCGAGCTCTGAACCGTCTCGACGAGATCGTGATCTCTGATCTCTGTCAC 2459
 2475 CAGTTGAGGAGGAGTGTGCTGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2534
 2460 GCAGCTGAGGAGGAGTGTGCTGCTGACGAGGAGGAGGAGGAGGAGGAGGAGG 2519
 2535 AGGAGTGTCTTTCGAGTCACTGATGCTGTTTGGACTATATCTTGGCAGAGGATG 2594
 2520 GGGATCGCTCTGCTGTCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2579
 2595 CCGGCTGTATGTTGCTGAGGCTTATAGGAGATGAGTGGAGAGGAGGAGGAGGAGG 2654
 2580 TCCGCTGTACCGGCGCGGCGCAATCAGGAGTGTGATCGAGAGGAGGAGGAGGAGG 2639
 2655 GTCAAGAGTGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2710
 2640 GTCAAGAGTGTGATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2699
 2711 --CAGCGCTGCTGATCTGTTGACCGGCTAGA---AAGTGGAGGAGTGTGAGGAGG 2765
 2700 GCGCGCAGGAGGAGGAGTGTCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2759
 2766 AACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2825
 2760 GACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2819
 2826 AGCTCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2882
 2820 TGGCGAGGCGGCTCAAGAGGATGAGGATCATGGAGGAGGAGGAGGAGGAGGAGGAGG 2876

SUPL 10
 174433
 CUS
 Trifolium aestivum heat shock protein 101 (HSP101c) mRNA, complete cds.
 Trifolium aestivum (bread wheat)
 Trifolium aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.
 1 (bases 1 to 3058)

AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source
 gene
 CDS
 BASE COUNT
 ORIGIN
 Query Match
 Best Local Similarity
 Matches 1887; Conservative 0; Mismatches 803; Indels 21; Gaps 5;
 QY 163 ATGAATCCGAGAAATTCACACACAGCAACAAAGAGCAATTCACAGCTCATGAGCTA 222
 Db 94 ATGAATCCGAGAAATTCACACACAGCAACAAAGAGCAATTCAGCGCTCGCGCGGCG 153
 QY 223 GCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 282
 Db 154 GCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 213
 QY 283 GATCCCGCGGTATATTTCTTCAAGCAATCTTAGTGC-----CGTGGCGAGCAACGCA 336
 Db 214 GACAGGTCGGGATCTCTCCGCGAGGAGGATGATGATGATGATGATGATGATGATGAT 273
 QY 337 GCTCAATGCTGTAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 396
 Db 274 GCGAGTGTGTCGAGGCGGTCGCGTCCGCGGCGTCAAGCGGCTGCGCTCGCAGTCCCCG 333
 QY 397 CCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 456
 Db 334 CCGCCGAGACCGGTCGCGGCGCTCCACCTCGCTGATGATGATGATGATGATGATGATGAT 393
 QY 457 GCTCAGAGGTCAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 516

Campbell,J.L., Klueva,N.Y., Zheng,H.G., Nieto-Sotelo,J., Ho,T.D.
 and Nguyen,H.T.
 Cloning of new members of heat shock protein HSP101 gene family in
 wheat (Triticum aestivum (L.) Moench) inducible by heat,
 dehydration, and ABA(1)
 Biochim. Biophys. Acta 1517 (2), 270-277 (2001)
 21240210
 11342108
 2 (bases 1 to 3058)
 Klueva,N.Y., Campbell,J.L., Zheng,H. and Nguyen,H.T.
 Direct Submission
 Submitted (02-AUG-1999) Plant and Soil Sciences, Texas Tech
 University, Mail Stop 2122, Lubbock, TX 79409, USA
 Location/Qualifiers
 1. 3058
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultiVar="Mustang"
 /db_xref="taxon:4565"
 1. 3058
 /gene="HSP101c"
 94. 2835
 /gene="HSP101c"
 /note="ClpB family member"
 /codon_start=1
 /product="heat shock protein 101"
 /protein_id="AA01280.1"
 /db_xref="GI:6013196"
 /translation="MNPDKFTHTNEALAAAHMAGHAQLPLHLAAALAAADRSG
 ILRCIAHAGSGNDAAAESEFVARSALKEPLPSQSPDPTVPASTSLVKAVERASQAO
 KSRGDSHADVQLMGLEDDPQISDALKEAGISAAARVKAEEVKLRGDNRRVESAGD
 TNFQALTYGRDLVAGKLDVIRDEEIRRVVRLIRKNNPVLIGPFGVKGATV
 VEGLAQRVWDGVPDNLVLDVALMGALVAGAKYGEFEERLKAIVLKEVEAEKGV
 ILFDEIHLVIGARTGSDMAANLFKMLARGQLRCIGATTLLEKYVEKDAEER
 RFQGFVARESPVPTVSLRLEKEVEGHGVRIQDRALVIAAQLSRYITMGHLPDK
 AIDLVACANRVOLDSPPEIDNLERKRIOLELHLEKEDKASKARLVDRVKE
 LDDLRLKLPLOMKYRKEKEIDEIRSLKORRELOFTLOEARRMDLARVALRYGA
 LQEVDAIAIKBETGENTMLTETGPDQAEVVSRTWGTPTVRLQNEKARLIGLAD
 RLHVRVVGVEAVNAVEALVRSRAGLRPQPTGSLFGLPTGVGTGKTEALAKASQL
 FDDNLVLRIDMSEYHSHVRLIGAPPYGVHGGGQLTEQVRRPYSVILFDEVE
 KARHVAVENTLQVLDGRLTDGQRTVDPRNTVIMTSLGAHLLAGMVGNSMKVAR
 DLWQGVRSRHRPELLRLDEIVIFDPLSEQLRKVARLQMDVAVLAERGVAVLT
 DLAQVDSLSYDPVYGARPIRMIEKRIKRVITELSKMLIREIDENSTVIDAAPSDE
 LITVDKHGSLVNARTGHKSDILIQVPSVAVGDAHAHVKKIMQDSGEVDDMEEE"

394 GCGCAGAAAGTCGCGGGGACCTCGACCTCGCGCTCGACCAAGCTGCTCATGGGCTCTCTC 453
517 GAAGATTCTCAATCAGGGAATTTGTTGAACGAGTTCGGTGTAGCAGCCGCGAGGTTAAAG 576
454 GAGGACCCGCGAATCTCCACCGCTCAAGAGGCGCGCATCTCCGCTGCGCGGTGAAG 513
577 TCTGAGGTTGAGAAGCTTCGTGGGAAAGAGGAAGTTGAGAGTGTCTTCAGGGAC 636
514 GCGGAGTTCAGAGCTCCGGGAGGCGACACCGCGCGTGGAGTCCGCTCGGGGAC 573
637 ACAATTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGGAAGCTT 696
574 ACCAATCTCCAGGCCCTCAAGACGTACCGCGCGACCTCGTGAGAGTGGCGGCAAGCTG 633
697 GATCTGTGATTTGCTGCTGATCAGGAGATTAGAGTGTGAGAGTCTTTTCGAGGAGA 756
634 GACCGGTCTATCGCCCGCAGCAGGAGATCCGCGCGTGGTCCGATCTGTCGCGCGC 693
757 ACGAAGAACAACTCTGTGCTTATTTGAGAGCCAGGAGTTGGTAAACAGCTGTGTTGAA 816
694 ACAAGAAACAAACCCGCTCTCATCGCGAGCCCGGTGGGCAAGCCCGTGTGGAG 753
817 GGTTAGCAAAAGGATTTGAAAGAGATGTGCCCAACAGCTCTTACTGATGTGAGATTA 876
754 GGGCTCGCGCAGCGGTCTGCGCGCGACGCTCCCGACCACTCTCTGAGCGTGGCGCTG 813
877 ATTTGCTTGCACATGGTTCGCTTAGTGTGCTGCTTAATACCGAGGAGGTTTGAAGAA 936
814 GTCCGCTCGACATGGGCGCTCTGTGGCGCGCCAGTACCGCGCGAGTTCGAGGAG 873
937 AGGTTGAAATCTGTTTGAAGAAGTTGAGGACGCTGGAAGGCAAAAGTGAATCTCTTTAT 996
874 CGGCTCAAGCGCTGCTCAAGAGGTTGAGGAGCGCGGTTCCGAGGCGGTGATCTGTTATC 933
997 GATGAGATTCATTTGGTCTTGGTGTGCAAACTGAAGGCTGATGATGACGCTAAT 1056
934 GACGAGATACACTGCTGTGCTCGCGCGCGGAGCGAGGGGTCAATGACGCGCGCAAC 993
1057 CTGTTCAAGCCCATGTTAGTACGAGGCGAGCTTCGATGCTATGTTGCTACACGCTTGA 1116
994 CTGTTCAAGCCCATGTTGCGAGGGGCGAGCTCAGGTGATTTGGCGGACGACCTCGAG 1053
1117 GAATACAGGAATATTTGAGAAAGATGCTGCTTTGAGAGGAGGTTCCAAACAGTCTAT 1176
1054 GAGTACAGGAAGTACGTCGAGAAGACGCGGTTTCGAGAGCGGTTCACGAGGTTC 1113
1177 GTTGGGAGCAGTGTGCTGACACCATTAGTATCTTAGGAGCTCAAGGAGAGTAT 1236
1114 GTGGCGGAGCCAGCGTCCCGACACCGTCACTACTCAGAGGCTCAAGGAGAGTAC 1173
1237 GAGGACATCATGTTGCGAATCCAAGACAGAGCTTTATAATGCTGCTCAGCTGCT 1296
1174 GAGGGGACCATGGGTGAGGATTCAGACCGCGGCTCGTCAITTTGCCGCGAGCTTCG 1233
1297 GCTGTTACATAACTGGTGGGCAATTTACGGATAAGCAATTTGATTTGGTTGATGAGGCT 1356
1234 TCGAGTACATCATGGTGGCCATTTGCTGTATAAGCAATCGATTTGTTGAGGAGGCC 1293
1357 TGTGCAATGTGAGAGTCAGCTTGAATGATCACTGAGAGATGATTAACCTTGAAGG 1416
1294 TCGCAATGTGAGGTTGCACTTTCAGCAGCCGCTGAAGAGATCGACCACTCGAGAGG 1353
1417 AAGAGATCAGCTGGAATTTGAATCTTACGCTTTGAAAGGAGGAGATTAAGCCAGC 1476
1354 AAGAGATCCAGTTGAGTTCAGCTTCAATGCTTCGAGAGGAGAGGACAGGCTAGC 1413
1477 AAGCTCGACTTATAGAGTGGGAAAGAGTGTGATGACCTGAGAGCAAGCTTCAGCT 1536
1414 AAGCTCGCTAGTTGATGTGAGGAAAGAGTTGGACGATCTGAGAGACAAGCTGAGCG 1473
1537 CTCAGATGAATACAGAAAGAGAGAGAGATTTGATGATTCGAGGCTTAAACAG 1596
1474 CTGCAGATGAGTATCGCAAGGAGAGAGGATTCGACGAGATCAGGAGCTTGAAGCAG 1533

1597 AAAAGAGAGAGCTCATGTTTCTTTGAGAGGAGCAAGAGATATGACCTTTGCAAGA 1656
1534 CGCCGCAAGAGCTGCAATTTCTCTGAGAGGCGCGAGCGCGGATTTGGCCGC 1593
1657 GCTCTCATCTAAGATATGCGCAATTTCAAGAAAGTGAATCTGCAATTTGCCCACTTGA 1716
1594 GTGGCTGATCTCAGATACGTTGCTCTCAGAGAGGTTGACGCTGCTTGTAACTGAG 1653
1717 GBAATCTTCTTGAAGAGAAATGTGATGCTCAAGAAACGTTTGGGCTGAAACATTTCT 1776
1654 GG---TGAGACCGCGGAGAAATTTGATGTTAAAGAGACTGTGCGCGCGGACCATTTCT 1710
1777 GAGGTTGTGAGCGGTTGCAAGGATTTCCAGTACGAGACTTTGGCCAAATGTAGAGAG 1836
1711 GAGTGTGTGAGCGCTGACTGTTGTTCTGTTCCAGGCTCGGACAGAAATGAGAGGCG 1770
1837 AGGTTGATTTGCTCTTCTGATAGGTTGCTAAGCGGTTTGGGACAGAAATCAAGCGGTA 1896
1771 AGGCTGATCGGCTGGCAGATCGACTGATCAGAGGTTGTTGACAGTATGAGGCGCTC 1830
1897 AATGCAATTTCTGAGGCAATTTCTAAGTCAAGGCGAGAACTTTGGTAGGGCAACAGCCA 1956
1831 AATGCAATTTCTGAGAGGCTGTTCTACGTTGAGGCGCTGGCTTTGGAAGCGCGCAGCCT 1890
1957 ACTGATCATTTCTTATTTCTTGGACCAACTGTTGGCAAACTGAGCTCGCAAGGCT 2016
1891 ACTGTTCTATTTCTTCTTGGACCGAGCGGTTGCAAAACCGAGCTCGCAAGGCT 1950
2017 CTGCTGAGCAGCTGTTTGTATGATGAAACCTCTTATGTTGCGATTTGATATGCGAATAT 2076
1951 CTAGCTGAGCAGCTGTTTGTATGAGAGAACTCTGCTCGCATCGACATGCTGTAATAC 2010
2077 ATGGAACAACTCTGCTCTCGCTCAATGGGCGACCAAGGCTATGTTGCTACAGG 2136
2011 ATGAGCAGCATTTGCTGCTGCGGCTAAATCGAGCGCCCACTCGGATATGTTGCTCATGAA 2070
2137 GAAGTGTGACAACTAATCTGAGGCTGTGAGGAGCGACCTTATTTGCTCATCTTTGAT 2196
2071 GAAGGGGGGAGCTGACCGAGCAAGTGAAGAGAGGAGGCAATACAGTGTCTCTTTGAC 2130
2197 GAAGTGTGAGAGGCTCATGTTGCTGCTTCAACACTCTGCTCCAAAGTTTGGATGATGTT 2256
2131 GAGTTGAGAAAGCGCATGTTGGCGGTTTCAACACTCTGCTCCAGGCTCTTGGACGATGG 2190
2257 CGATTTGACAGCGGCGCAAGGCGAGCAGTCTGATTTCAAGAACTCGGCTGATATCATGACA 2316
2191 CGGTTGACCGAGCGGCGAAGGCGAGCGGTTGATTTTGAAGACACGCTGATCATGATGCC 2250
2317 TCAAACTTGGTCTGTAACACTCTTTCAGGCGCTAACTGGGAAAGTAACTTGAAGTG 2376
2251 TCAAACTTGGCGCGGAGCACTCTTCGCGGATGGTGGGCA---TTTCGATGAAGTT 2307
2377 GCCCGGAGCTGTGTGATGCGGAGGTGAGGAAACACTTTCAGACCAAGAGCTCTTGAACAG 2436
2308 GCTGCTGATCTGCTGATGAGGAGGTGAGGAGCAATTTCCGCGCGGAGCTGCTGAACCT 2367
2437 CTTTGAAGAGATTTGTTGTTGCGCCCTTTTCACTGACAGTTCAGGAAAGTACGCTCGG 2496
2368 CTGGAAGAGATCTGCTCTTTCGACCTCTGTCGATGAGCAGCTGCGAAGGCTCGCTCGG 2427
2497 CTTTCAATGAAGACGCTGCTGTCGCGCTTGTGAAGAGGAGTGTCTTTTGGCAGTCACT 2556
2428 CTTTCAATGAAGATTTGCGAGTCTGCTTTCGAGAGGCGGCTGCTCTGCGCGCTCAC 2487
2557 GATGCTGCTTTGAGCTATATCTTTGGCAGAGATTTATGACCCCGTGTATGCTGCTAGGCT 2616
2488 GACGCGCTTGGAGCTCATCTGCTCACTGCTTTACGATCCGCTCTATGCGCGCAGGCA 2547
2617 ATGAGGAGATGATGAGAGAGAGGTTGAGAGAACTCTCAAGATGCTTCAAGTGTGCTGAG 2676
2548 ATCCGAGATGGATCGAAGAGAGGATGATGAGCGAGCTCTCAAGATGTTGATCCGCGAG 2607

2677 GAAATCGATGAAATCCCACTGTTTACATAGATGCAG-----GGCTGTGTGATCTTGTG 2730
 2608 GAGATCGACGAGAACTCCACGGTGTATCATCGACGCTGCGCCCAAGCAAGGACGAGCTGACC 2667
 2731 TACCGGTAGAAA---GTGAGGTCTAGTGGAGCTTCAACAGCGAAGAGTCAAGATG 2787
 2668 TATGGCGTGCACAGCAAGCAGGAGGCTGGTGAACGCGCGCACCGGCCCAAGTCCGACATC 2727
 2788 CTGATTCATATTCGTAACCGCGCCAAAGAGAGTGTGACGCTCAGCGCGGTGAAGAGATG 2847
 2728 CTGATTCAGGTTCTTACGCGAGCTGTGTGGGGGCGATGCGCGCGACCGCTGAAGAGATG 2787
 2848 AGGATCGAGGA 2858
 2788 AAGATCATGCA 2798

SULT 11
 332981
 CUS
 AF332981 3148 bp mRNA linear PLN 02-JAN-2002
 Oryza sativa heat shock protein (HSP101) mRNA, complete cds.
 AF332981
 AF332981.1 GI:18033449
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoideae; Oryza.
 1 (bases 1 to 3148)
 Agarwal,M., Sahi,C., Young,T., Agarwal,S., Mitra,S., Ganesan,K.,
 Agarwal-Katiyar,S., Gallie,D.R. and Grover,A.
 Isolation and Analysis of HSP101 from Rice
 Unpublished
 2 (bases 1 to 3148)
 Young,T.E. and Gallie,D.R.
 Direct Submission
 Submitted (02-JAN-2001) Biochemistry, University of California,
 Boyce Hall, Riverside, CA 92521-0129, USA
 Location/Qualifiers
 1. 3148
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 1. 3148
 /gene="HSP101"
 135..2873
 /gene="HSP101"
 /codon_start=1
 /product="heat shock protein"
 /protein_id="AAL57165.1"
 /db_xref="GI:18033450"
 translations="MNPDPNFTNEALVAHEILASEGHAQLTPLHVAALADKGG
 ILRAISQASGGDAGPADPSPFVSVGALKLPSQPPSPVSPASTALIKVIRASQ
 KRGDSHLAVDQLGLLESLDCLKAGVSAARVRAELERKRGGRKVESASGD
 TNPQALTKYRDVLDVLRALDQGLVAGAKYRGEFEERLKAFLKEVEAEAGKV
 VEGALQIRVGRDVPVSLDVLRLDQGLVAGAKYRGEFEERLKAFLKEVEAEAGKV
 ILFDIHLVLAGRGTEGMDAANLKPMLARGQLRCIGATLLEVKYVKEDAAEP
 RFOQVFAEVSVPDTISILRGLKEKVEGHGVRIODRALVVAALASARYIMGRHLDPK
 AIDLVEACANVRVQDSPEEDINLERKRIQLEVEHHALEKEKAKARLVEYKE
 LDRLDKLQPLTKMYEKERIDEIKLQKREELQFTQAEERRMDLARVADLKTA
 LQEDIVAIKQSETEGKMLTETVPEQAEVSVSWTGIPTVRLQNDKERLVGLAD
 RLHQRVQAEAVSAVAELRAGLRPQPTGSLFGLPTGVKTEKALAEQL
 FDDENLLVIRDMSEYMEQHSVARIKAPGVYGHGEGQLTEQVRRPYSVILFDEVE
 KAHVAVENTLLQVLDLGRITDGGRTVDENTVIINTSLGAEHLILGAMVGNKMKVA
 RDLVMEVRRHPRFLNRLDELVIIFDPLSHEQLKVARLQMDVAVRLAERGVLA
 TDAALDVLSLSYDPVYGTGPRIRWLEKRVVITOLSQMLIQBELDENCTVIIDAAHPKO
 ELAYRVDNRGLVNAETGQKSDILIQVNPNGAATGSDAAQAVKMKRIMDEDDGMDSE"

source
 gene
 CDS
 153 ATGATTCAGAGAAATTCACACACAGCAACAGCAAAATTCAGCTCATGAGCTA 222
 135 ATGAATCCGGACAACTTTCACGCAACAGCAACAGCGCGCTCGTGGCGGCGCA 194
 223 GCTGTGAATCAGAGACATGCTCAATTCCTCTTTGATTTAGCTGTGCTTTGATCTCT 282
 135 GCGTCGAGGCGGGGCGCGCAGCTGACGCGCTGACCTGTGTGGCGGCGCTGGCGGCG 254
 283 GATCCCAACCGGTATATTTCTCAAGCAATCTCTAGTCCGGTGGCGGAGAAACGAG----- 337
 255 GACAAAGGCGGGGATCTCCGGCAGCGATCTCGCAGCGCTCCGGCGGGGAGCGCGGCGG 314
 338 -CTCAATCTGCTGAAAGAGTCAATCAATCAAGCCCTTGAAGAGCTTCTTCAATCTCTCT 396
 315 CCGGACTCGTTCGAGCGCGCTGTGAGCGGGGCGCTCAAGAAAGCTGCCCTCGCAGCTGCGG 374
 397 CCACCTGATGATATTTCCAGCGAGTTCCTAGTCTTATTAAAGGTCAATTCGTCGTGCTCAAGCT 456
 375 CCGCGGACAGCGTGGCGGCTCCACGCGCTGATCAAGGTCAATCCGCGCGCGCAGTCC 434
 457 GCTCAGAACTCAGAGGTGATCTCATTTGGCTGTTGACCAAGTTGATATGGTCTTCTT 516
 435 GCGCAGAAAGAGCGCGGGGACTCCCACTCCGCTCGACCAAGCTGCTGCTGGGCTCTCTG 494
 517 GAAATCTCTCAATCAAGGATTTGTAACGAAGTCTGTAGCGAGCGGCGAGGTAAG 576
 495 GAGGACTCGCTATCTCCGACTGCTCAGAGAGCGCGGCTGTCGCGGGCGGGTCCGC 554
 577 TCTGAGGTGAGAGCTTCTGTGGGAAAGAGGAGAAAGTTGAGAGTGTTCAGGGGAC 636
 555 GCGGAGCTCGAGAGCTCCGTCGCGGGGAGGCGCAAGGTGAGTCCGCTCGCGCGAC 614
 637 ACAATTTTCAAGCTTTAAGACTTATGGAAGACTTTGTTGAGCAAGCAGGAGAGCTT 696
 615 ACCAACTTCCAGGCGCTCAAGACCTACGCGCGGACCTCGTCGAGCAGGCGCGTAAGCTC 674
 697 GATCTGTGATTTGCTGTGATGAGAGATTAAGAGTCTGTGAGGATTTCTTTCAGGAGA 756
 675 GACCCGCTCATTTGCGCGGCGAGAGAGATCCGCGCGCTGTGTGCGCATCTGTGCGCGCG 734
 757 ACGAAGAACTTCTGTGCTTTATTTGAGAGCAGAGAGTGTGTAACAGCTGTGTTGAA 816
 735 ACGAAGAACTTCTGTGCTTTATTTGAGAGCAGAGAGTGTGTAACAGCTGTGTTGAA 794
 817 GGTTTAGCACAAGAGTTGTGAAGAGAGTGTGCGCAACAGCTCTTACTGATGTGATTA 876
 795 GGCCTCGCGCAGCGCATCTGTGCGGGGAGCGTCCCGAGCAACCTGCTCGAGCTCCGCTC 854
 877 ATTTCTGTTGACATGGGTGCGTTAGTTGCTGCTAAATACCGAGGAGAGTTCGAAGAA 936
 855 ATCCGCTCGACATGGGCGCGCTCGTCGCGCGGCGCAAGTACCGCGCGAGTTCGAAGAA 914
 937 AGTTGAAATCTGTTTGAAGAGAGTGTGAGAGCAGCTGAGAGCAAGTGTCTTTATT 996
 915 CGGCTCAAGGCGGCTGCTCAAGAGGAGTGTGAGAGCAGGAGGAGTGTCTCTTCTATC 974
 997 GATCAGATTCATTTGCTGCTGCTGGCAAACTCAAGGCTCGATGATGAGCTAAT 1056
 975 GACGAGATACCTTCGCTCGCGCGGAGGAGCGAGGCTCCATGGAGCGCGGCAAC 1034
 1057 CTGTTCAAGCCCATGTTAGCTAGAGGCGAGCTTCGATGCTTGTGCTACACGCTTGAA 1116
 1035 CTGTTCAAGCAATGCTCGCGAGGCGGCGAGCTCCGCTGCTCGCGCGCACGAGCTGGAG 1094
 1117 GAATACAGAAATATGTGAGAAAGATGCTGCTCTTTAGAGGAGGTTTCAACAGCTAT 1176
 1095 GAGTACCGCAAGTACGTCGAGAGGAGCGCGCTTTCAGAGGCGGTTCCAGCAGGCTTC 1154
 1177 GTTCGGGAGCAAGTGTGCTGACACCATAGTATCTCTTAGAGGACTCAAGGAGAGTAT 1236

Query Match

41.5%; Score 1289.2; DB 8; Length 3148;

SE COUNT
 IGIN

644 a 931 c 1087 g 486 t

2317	TC	AA	AC	CT	TG	TG	CT	GA	CA	CT	CT	TG	CA	GG	GC	T	A	CT	GG	AA	GT	A	CA	A	T	GG	A	AG	TG	2376					
2292	TC	CA	AC	CT	CG	CC	CG	CA	CT	CT	CG	CG	CA	TG	TG	CG	CA	AG	AA	CT	CA	T	A	GA	AG	GC	2351								
2377	GC	CC	GG	AG	CT	GT	GA	TG	CG	GG	AG	GT	G	A	A	A	C	A	CT	T	C	A	G	A	CT	T	T	G	A	C	AG	2436			
2352	GC	CC	GT	GA	CT	CT	CA	TG	CA	GG	AG	GT	G	A	G	A	C	A	CT	T	C	CG	T	C	T	G	AG	CT	T	C	GA	CC	GG	2411	
2437	CT	T	G	A	C	G	A	G	A	T	T	G	T	T	G	C	A	C	C	C	T	T	T	C	A	T	G	A	C	C	A	T	G	2496	
2412	CT	C	G	A	C	G	A	T	T	G	A	T	T	G	A	T	T	C	C	C	T	G	T	C	C	A	C	A	G	A	G	T	C	2471	
2497	CT	T	C	A	A	T	G	A	A	G	A	C	T	T	G	T	C	G	C	G	T	T	G	A	A	G	A	G	A	G	T	C	T	2556	
2472	CT	C	A	G	A	T	A	A	G	A	C	T	C	G	C	G	T	C	G	C	G	C	T	C	G	C	G	A	G	A	G	T	C	2531	
2557	G	A	T	G	C	T	G	C	T	T	G	A	C	T	A	T	C	T	T	G	G	C	A	G	A	G	A	G	T	T	A	G	A	C	2616
2532	G	A	C	C	G	C	G	T	G	A	C	T	C	T	G	T	C	T	C	T	A	T	G	A	T	C	C	G	G	T	T	A	T	C	2591
2617	A	T	A	A	G	A	G	A	T	G	A	A	G	A	G	T	G	T	G	A	C	A	A	A	C	T	G	T	C	A	A	A	G	T	2676
2592	A	T	C	A	G	A	G	T	G	A	T	C	A	A	G	A	G	T	G	T	G	A	C	A	A	G	A	G	T	C	A	A	G	A	2651
2677	G	A	A	T	C	G	A	T	C	C	A	T	C	T	G	T	T	A	C	A	T	A	G	A	T	G	C	A	G	A	T	T	G	2730	
2652	G	A	G	A	T	C	G	A	A	C	T	G	A	C	C	G	T	G	A	C	A	T	C	G	A	C	G	C	G	C	C	A	A	G	2711
2731	T	A	C	C	G	G	T	A	G	A	---	---	G	T	G	A	G	A	G	T	C	A	G	G	T	T	C	A	A	G	A	A	G	T	2787
2712	T	A	C	A	G	G	T	G	A	C	A	C	C	G	G	G	G	G	T	T	G	T	G	A	A	C	C	G	C	G	A	A	G	T	2771
2788	C	T	G	A	T	C	A	T	T	G	C	T	A	A	C	G	---	---	G	C	A	A	A	G	A	G	T	G	A	T	G	C	A	G	2844
2772	C	T	C	A	T	C	C	A	G	G	T	C	C	C	C	C	A	C	C	G	G	A	C	C	C	G	G	A	C	C	C	G	A	A	2831
2845	A	T	G	A	G	A	T	A	T	A	G	A	G	A	T	A	T	A	G	A	G	A	T	A	T	A	G	A	G	A	T	A	T	A	2892
2832	A	T	G	A	G	A	T	C	A	T	G	A	G	A	C	G	G	A	C	G	G	A	C	G	G	A	C	G	A	T	A	T	A	T	2869

AF097363 3107 bp mRNA linear PLN 14-MAY-2002
Triticum aestivum heat shock protein 101 (Hsp101b) mRNA, complete
cds.

AF097363
AF097363.1 GI:4558483

Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Triticum.

1 (bases 1 to 3107)
Campbell, J.L.; Klueve, N.Y.; Zheng, H.G.; Nieto-Sotelo, J.; Ho, T.D.
and Nguyen, H.T.
Cloning of new members of heat shock protein Hsp101 gene family in
wheat (Triticum aestivum (L.) Moench) inducible by heat,
dehydration, and ABA(1)
Biochim. Biophys. Acta 1517 (2), 270-277 (2001)
21240210
11342108

2 (bases 1 to 3107)
Campbell, J.L.; Klueve, N.Y. and Nguyen, H.T.
Direct Submission
Submitted (05-OCT-1998) Plant and Soil Science, Texas Tech
University, Mail Stop 2122, Lubbock, TX 79409, USA
Location/Qualifiers
1. 3107
/organism="Triticum aestivum"
/mol_type="mRNA"

694 TTGTATCTCTGTGATTCGGTTCGTGATGAGGAGATTGAAGAGTCTGTGAGGATTTCTTTCGAGG 753

652 CTGACCCCGTCTATTCGGCCGCGACGAGGAGATTCGCGCGCTGTGCGCATCTCTCAACG 711

754 AGAACGAAGAACAACTCTGTCTTATTGGAGAGCCAGAGATTGGTAAACAGCTGTGGTT 813

712 CGCACCAAGAACAAACCCCGTCTCATCGCGAGCCCGCGCTGGGCAAGACGCGCGTGTG 771

814 GAGGTTTTAGCACAAAAGGATTTGTAAAGAGAGATGTGCCAACAAGCTTTTACTGTATGTGAGA 873

772 GAGGGCTTCGCGCAGCGCGCTCGTGTGCGGGGACGTGTGCCAGCAACTCTCTCGAGCTCGCG 831

874 TTAATTTTCTGTTGACATGSGTGGTTAGTTTGTCTGTCTGCTTAATAACCGAGGAGAGTTTGAA 933

832 CTGTCGCGCTCGACATGGCGCGCTCTGTGCGCGCGCAAGTACCCGCGCGAGTTCCAG 891

934 GAAAGGTTGAAATTCGTCTTTTGAAGAGATTGAGGACGCTGAAGGCAAAAGTGAATTCCTTTT 993

892 GAGCGGCTCAAGGCCGTGTCTCAAGGAGTGGAGGAGGCCGAGGGCAAGGTCATCTCTTC 951

994 ATTGATGAGATTCATTTGGTTCTTGGTGTGSCAAAACCTGAAGGTCGATGGATGCAAGCT 1053

952 ATCGACAGATACCACTCTGTCTCGCGCGCGGGAGGAGCGAGGGGTCCATGGACGCGGCC 1011

1054 AATCTGTTCAAGCCCATGTTAGCTAGAGGGCAGCTTCGATGCAATTTGGTGTCTACAACGCTT 1113

1012 AACCTGTTCAAGCCCGATGCTCGCCAGGGCCAGCTCCGGTGCATCGCGCCACCACGCTG 1071

1114 GAGAATPACAGGAAAATATGTTGAGAAAGATGTGCTTTTGAGAGAGAGGTTCCAAACAATGC 1173

11072 GAGGAGTACCGCAAGTACGTTGAGAGAGACGCGCGGTTTCGAGCGCGGTTCCAGCAGGTG 1131

1174 TATGTTCCGGAGCCAAAGTGTGCTGACACCATTAATGATTCCTTAGAGGACTCAAGGAGAAG 1233

11132 TAGGTGCGGAGCCGAGCGTGTGCGCGACCATCAGCATCTCTCGAGGGCTCAAGAGAG 1191

1234 TATGAGGGAATCATCGTGTGCGAATCCAAAGACAGAGCTTTATAAATGTGCTCAGCTG 1293

11192 TAGAGGGGACACACGGCTGCGCATCCAGGACCGCGCATCGTCTCGCGCGCAGCTC 1251

1294 TCTGCTGTTTACATAACTGTTGGCATTTTACCGGATTAAGCAATTGATTTGGTTGATCAG 1353

1252 TCGCGAGGTTACATCATTTGGCGCGGACCTGCGCGAACAAGGCCATCGACTCGTGGAGCAG 1311

1354 GCTTGTCCGAATGTGAGGTCAGCTTGATGTAGCTCAACTGAAGAGATTTGAATACTTGAA 1413

1312 GCGTGCACAACTGAGGTTGACGTGAGACCGCAGCCGAGGAGATCGACAACCTGGAG 1371

1414 AGGAAGAGGATGACGTGGAAATTGAATTCACGCTTGGAAAGGGAGAGAGATTAAGCC 1473

1372 AGGAAGCGGATTCAGCTGGAGTGGAGCTGCAAGCGCTGGAGAGGAGCAAGGCAAGGCC 1431

1474 AGCAAAAGCTCGATTTATAGAGTTCGGAAGAGCTTGATGACCTGAGAGACAAGCTTCAG 1533

1432 AGCAAGCCCGGCTGTGTGAGGTGAGGAAGAGAGCTGAGACCTGAGGGACAAGCTGCAG 1491

1534 CCTCTCAGATGAATAACAAAGGAGAAAGAGAAATTGATGAGATTCGAAGGCTTAAA 1593

1492 CCGCTGACGATGAAGTACAGGAAGGAGAGGAGCGGATCGACAGATCCCGAAAGCTGAAG 1551

1594 CAGAAAAGAGAGAGCTCATGTTTTCTTTTTCAGGAGGACGAACGAGNATATGACTTTGCA 1653

1552 CAGCGGCGAGAGAGCTGAGTTTCAAGCTGCAAGGAGGCGGAGCGCGCATGGAACCTGGCC 1611

1654 AGAGCTGCTGATCTAAGATATGCGGCAATTTCAAGAAGTGGAACTTGCAATTTGCCCAACTT 1713

1612 GCGCTGCGCGACCTCAAGTACGGCGCTCTCAGGAGATCGACCGCGCATCGCCAAGCTG 1671

1714 GAAGGAATCTTTCTGAAGAGAAATGTGATGCTCAACAGAAAACGTTTGGGCTTGAAACAATT 1773

1672 GAGG---CGAGACCGGCGAGAACCTGATGCTTCAACGAGACCGTTCGGCGCGAGCAGATC 1728

1774 GCTTACGCTCTGAGCGCGTTGCAACGCGATTCCTACCTGAGCGAGCTTGGCCCAAAATGAGAAG 1833

Query Match	40.2%; Score 1246.8; DB 8; Length 3107;
Best Local Similarity	67.6%; Pred. No. 2.9e-304;
Matches 1851; Conservative	0; Mismatches 857; Indels 30; Gaps 6
163	ATGAATCCAGAGAAATTACACACAAGACAACAGACAAATTTGTCAGAGTCATAGACTA 1222
112	ATGNACCCGGACAACTTTACGCCACAAAGACCAACAGAGCGCTGTGTGGCGGCGCAGAGCG 171
223	GCTGTGAATTCAGGACATGCTCAAATTCATCTCTTTTGCATTTTAGCTGTGTCTTTGATCTCT 2823
172	GCGTCGAGGCGGCCACGCGCAGATCAGCGCTGCACCTTGGCGGCGCGCTGGCGGCG 231
283	GATCCACCGGTATATTTCTCTCAAGAAATCTCTAGTGCGGTGGCGAGAA-----CGCA 3366
232	GACAAGTCGGGATCTGGCGCAGGCGCTCGCGGGGGCGTCCGGCGGGAACGCTCCGCG 291
337	GCTCAATCTGCTGAAGAGTGATCAATCAAGCCTTGAAGAGCTTCTCTTCACATCTCTCT 396
292	GGGAGTCGTGTCAGCGCGTGTCTCGCGGCGCTCAAGAGAGCTGCGTGCAGTCTCCCC 351
397	CCACCTCGATGATATCCAGCGAGTTCTAGTCTTATAAGGTCATTCTGTCGTCTCAAGCT 456
352	CCGCGGACTCTCGTCCGCGCTCCAGCGGCTCATCAAGGCCATCCGGCGCGCGCAGTCC 411
457	GCTCAGAAGTCAGAGTGATNACTCATTTTGGCTGTGTGACCAAGTTGATTATGGGTCCTTT 516
412	GCGCAGAAGAAGCGCGGGACTCCACCTCGCGCTCGACCAAGTGTCTATGGGCTCTCTC 471
517	GAAGATTCTCAATCAGGAGTTGTTGACGAAGTTCGGTGTAGCGACGGCGAGGGTAAAG 576
472	GAGACGCGCAGATCGCCACTGCTCTCAAGGAGGCGCGGTGTCCGCTCTCGCGGTGCGC 531
577	TCTGAGGTTGAGAAGCTTCGT---GGGAAAGAAGGGAAGTAAGTTGAGAGTGTCTCAGGG 633
532	GCGAGCTCGAAGAGTCCGCGCGGGGACAACCTCGCGAAGTTCAGTTCGCTCTCGGC 591
634	GACACAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTTTGGTTGACCAAGCAGGGAAG 693
592	GACACAACTTCCAGCGGCTTCAGAGTACGCGCGCGCACTGTCGAGTGTCGCGCGCAAG 651

1729 GCGAGGTGTGAGCGGTGAGACCGGCATCCCGTACCCCGCTCGGGGAGAACACAG 1788
1834 GAGAGTTGATTGGTCTTCTGATAGTTGATTAAGCGGGTTGTGGGACAGAAATCAAGCG 1893
1789 GAGCGCTCGTGGCATGCGGACAGCGTGCACACAGAGGTGTGGGGCAGACGAGGCC 1848
1894 GTAATGCGATTCTGAGCAATCTTAAGGTCAAGGCGAGGACTTGGTAGGGCACAACG 1953
1849 GTGAACGCGCTGCGGAGCGGTGTGAGGTCAAGGCGGGGCTTGGGCGCGCGAGCAG 1908
1954 CCAACTGAGTATCTTCTTATCTTGAACCACTGGTGTGGCAAACTGAGTCTCCCAAG 2013
1909 CCAACCGGTGTCTTCTTCTGTCGCGCCGACCGGTGGGCAAGCGAGTCTCCCAAG 1968
2014 GCTCTGCTGAGCAGCTGTTGATGATGAACCTTCTAGTTGCGGATGATATGCGGAA 2073
1969 GCGCTCGCGCAGCAGCTGTGACGACGAGAACCTGCTGTCGCGCATGACATGTCGAG 2028
2074 TATATGAAACACACTCTCTCTGCTCATGTTGGGACACACAGGCTATGTTGTCAC 2133
2029 TACATGGAGCAGTCTCTGCTGCGCGCTCATCGAGGCGCACCTGCTATGTCGGGCAC 2088
2134 GAGGAGGTGGACAACTAATCAGGCTGTGAGGCGGACCTTATTTGTCTATCTCTTT 2193
2089 GAGGAGGCGGCGAGCTGACGAGCAGGTGAGGAGGCGCGTACAGCTGATCTCTGTC 2148
2194 GATGAGTGGAGAGCTCATGTTGCTGCTTCAACACTCTCTCAAGTTTGTGATGAT 2253
2149 GACGAGTGGAGAGCGCACGTGGCGGTGTTCAACAGCTCTCTCCAGTGTGACGAC 2208
2254 GCTCATGACAGAGCGGCAAGCAGCAGCAGTCTGATTTTCAAGAACTCGTGTATCATG 2313
2209 GCGCGCTGACCGAGCGGCAAGCAGGAGCGGTGACTTCCGGAACACCGTGTATCATG 2268
2314 ACATCAACCTTGGTGTGAAACACTCTCTGAGGCGTAACTGGGAAAGTAAACATGAA 2373
2269 ACCTCAACCTCGCGCGGAGCACCTCTCGCGGATGGTGGGCAAGCACTCATGAG 2328
2374 GTGGCGCGGAGCTGTGATGCGGAGGTGAGGAAACACTTCAGACCAAGCTTTGAAAC 2433
2329 GTGCGCGGAGCTGCTCATGACGAGGTGAGGAGGCACTTCAGCGCGAGCTGTGAAAC 2388
2434 AGCGTTGACGAGATTGTGTTGTCAGCCCTTTCACATGACCACTTGGAGGAGTACT 2493
2389 CGCTGAGAGATGCTGCTTTCAGCCCGCTGCTCCAGCAGCAGCTGAGGAGGTGCT 2448
2494 CGGCTTCAAAATGAAAGAGCTGCTGCTCGGCTGTGTAAGAGGAGTGTCTTTGGCAGTC 2553
2449 CGCCTGCAGATGAAGAGCTGCGCGCTCGGCTCGCGAGCGCGGCTGCGCTGCGCTC 2508
2554 ACTGATGCTGCTTTGACTATATCTTGGCAGAGTATGACCCGGTGTATGTTGCTAGG 2613
2509 ACCGAGCTGCGCTGGAAGTCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2568
2614 CCTATAAGGAGATGATGAGAGAGGTGTGACAGAACTGTCAAGATGTTGTGCTGCT 2673
2569 CGATCAGAGGTGATCGAAGCGGTTGTTGACGAGCTGTCCAGATGCTGATTCAG 2628
2674 GAGGAAATCGATGAAATCTCACTGTTTACATAGATGAG ---GGCGTGGTGTATCTGTG 2730
2629 GAGGAGATCGACGAGAACTGACCGGTGTACATCGACCGCGCCCAACAGGACGAGTGGC 2688
2731 TACCGGTGAGA ---AAGTGGAGTCTAGTGGAGCTTCAACAGGCAAGAGTCTAGATGTG 2787
2689 TACCGGTGAGACCGGAGCGCGGCTGTGTGACCGCGAGAGCGGGCAGAGTCTGACATC 2748
2788 CTGATTCTATTGCTAACCGGCT ---AAAGAGAGTGTGACGCTCAGCGC 2835
2749 CTCATCCAGTCCCAACCGCGCTCTCGCGCGGCGCAAGCGGCGAGCGCGGCAAGGCC 2808
2836 GTGAAGAGATGAGGATCGAGGAAATAGAAATGACGA 2873

Db 2809 GTGAAGAAGATGCGGCTCATGAGGACGGGACGAGGA 2846

RESULT 13
AF083344
LOCUS
DEFINITION
Triticum aestivum heat shock protein 101 kDa (HSP101) mRNA,
complete cds.
ACCESSION
AF083344
VERSION
AF083344.2 GI:11561807
KEYWORDS
Triticum aestivum (bread wheat)
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE
1 (bases 1 to 2821)
Wells, D.R., Tanguay, R.L., Le H. and Gallie, D.R.
HSP101 functions as a specific translational regulatory protein
whose activity is regulated by nutrient status
Genes Dev. 12 (20), 3236-3251 (1998)
9784498
PUBMED
2 (bases 1 to 2821)
Gallie, D.R., Tanguay, R.L. and Wells, D.R.
Direct Submission
Submitted (10-AUG-1998) Biochemistry, University of
California-Riverside, Riverside, CA 92521-0129, USA
3 (bases 1 to 2821)
Gallie, D.R., Tanguay, R.L. and Wells, D.R.
Direct Submission
Submitted (06-DEC-2000) Biochemistry, University of
California-Riverside, Riverside, CA 92521-0129, USA
REMARK
Sequence update by submitter
On Dec 6, 2000 this sequence version replaced gi:3982832.
FEATURES
Location/Qualifiers
1..2821
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
1..2821
/gene="HSP101"
65..2821
/gene="HSP101"
/codon_start=1
/product="101 kDa heat shock protein"
/protein_id="AAC83689.2"
/db_xref="GI:11561808"
translation="MNPDNFTKTNALVAHEAASEAGHAQITPLHLAALAAKSG
ILROAVAGSGNAGSDSFEVLAGALRLPSQSPSPSPSPASTALIKAIRAQAQ
KRGDSHLAVQLMLGLLDAQIADCLKEAGVSASRVAELDKLRGDSNRKVESAFG
DTTFQALKTVGRDLVEVAGKLDPIVGRDEIRRVVRLSRRTNNPVLIGPVGKTA
VWGLAQRTVGRDVPNSLLDLVALDMGALVAGKYRGEFFERKAVLKEVEADGK
VILFIDEHLVAGTSGSDAANLFPKMLARGQLRCIGATTLEIRKYRVEKDAFE
RRFOVYVAPSVADTISILGKLEKRGHGVRIODRAIVVAOLSAFYIMGRHLPD
KADLVDEACANVRQLDSQPEIDNLEKRIQLEVELHALKEKDKASKARLVEVRK
ELDLKDLQPLTKRKERIDEIRKLQKREELQFTLQAEERRMDLARDVLKYG
ALQIEDAAIAKLEGETGENMLTETVPEQIAFVSRWTGIPVTRIGQNDKRLVGM
DLRHTVQVQTEAVNAVABAVLRSLAGLPQOQTFGLGFTGVGTTELAKALAEQ
LPDENALLVIDMSYVMEQSVARLIGAPGVVHEGGQLTEQVRRPYSVILFDEV
EKHAVENTLLQVLDGSLTDGQGTVDPRNTVIMTSLNCAHLIAGMVGKSNKY
ARDLVMEVRRHFRPELLNRLDENWIFDLSHEQLRKVARLQMKOVAVRLAERGVALA
VTDAALDILSLAYDPVYGARPIRWIEKRVVTLQSKMLIQEIDENCTVYIDAAKD
ELAYRVDKSGGLVNAETGQSDILITQVNGALGGGGGGEAAKAVKMKRMVEDGDEDG
DEDV"

BASE COUNT 524 a 902 c 1035 g 360 t
ORIGIN

Query Match 39.8%; Score 1236.2; DB 8; Length 2821;
Best local Similarity 67.3%; Pred. No. 1.4e-301;
Matches 1847; Conservative 0; Mismatches 868; Indels 30; Gaps 6;

QY 156 ATCGAAGATGAATCCAGAGAAATTCACACAGACAAACGACAGCAATTTGACAGCTCA 215

58 AGCGGCCATGAACCGGACAACTTCACGCAACCAAGACGAGCGCTGTGGCGGCGCA 117
216 TGAGCTAGCTGTGAATGAGGACATGCTCAATTCACCTCCCTTTGCTATTAGCTGTGCTTT 275
118 CGAGCGCGCTGCGAGCGCGGCGCACGCGGAGATCAGCGGCTTGACCTGCGCGGCGGCT 177
276 GATCTGTATCCCAACCGGTATATTTCTCAAGCAATCTTAGTSCCGGTGGCGAGAA --- 332
178 GCGCGCGCAAGTCCGCGATCTCTCGCGAGCGCTCGCGGCGGCGCTCCGCGGGAATGC 237
333 ---CGAGCTCAATCTGCTGAAGAGTATCAATCAAGCCCTTGAGAGCTTCTTTCACA 389
238 GTCGCGCGGGGACTCGTTGAGCGCGCTGCTCGCGCGCGCTCAGGAAGCTGCGGTGCA 297
390 ATCTCTCCACCTGATGATATTCAGCGAGTCTTAGTCTTATTAAGGTCAATTCGTCTGC 449
298 GTCGCGCGCGGACTCGTCCGTGCGGCTCCACGCGGCTCATCAAGGCCATTCGCGCGC 357
450 TCAAGCTGCTCAGAAAGTACAGAGTGATCTCATTTGCTGTTGACCAAGTTGATGGG 509
358 GCAGTCCGCGCGCAGAAAGACGCGGGGACTCGCACCTCGCGCTCGACAGCTGCTCATGGG 417
510 TCTTCTTCAGATCTCAATCAAGGATTTGTTGAACGAGTCCGTGTAGCGAGCGGAG 569
418 CCTCTCGAGACCGGAGATCGCGGACTGCTCAAGAGCGCGCGCTGTCGCGCTCGCG 477
570 GGTAAAGTCTGAGGTTGAGAACTTCGCT ---GGAAAGAGGAGAAAGTTGAGAGTGC 626
478 GGTGCGCGCGGAGCTCGACAGCTCCGCGCGGGGACAACTCGCGAAGGTTCGAGTCCGC 537
627 TTCAGGGGACAAATTTCAAGCTTTAAGACTTTGGAAGAGATTTGGTTGAGCAAGC 686
538 CTTGCGCGACACCACTTCCAGGCGCTCAAGACGTACGCGCGGACCTCGTTCAGAGTGGC 597
687 AGGAGCTTCATCTGCTGATGCTGCTGATGAGAGATTAGAGAGTCTGAGGATTCCT 746
598 CGGCAAGCTCGACCCCGCTATCGCGCGCGGAGAGATCGCGCGCTGTGCGGATCCT 657
747 TTCAGGAGAACGAAAGCAATCTGCTGCTTATTTGAGAGCGAGGTTGTTAAACAGC 806
658 CTCGCGCGCACCAAGAACACCCGCTCTCATCGCGAGCGCGGCTCGGCAAGCGC 717
807 TGTGGTTGAAGTTAGCAAAAGATTGTGAAGAGAGATGTGCCCAACAGTCTTACTGA 866
718 CGTCTCGAGGCGCTCGCGCAGCGCTGCTGCGCGGGGAGCTGCCCAAGCAACCTGCTCGA 777
867 TGTGAGATTAATTTCTGTTGACATGGGTGGCTTAGTTGCTGCTAAATACCGAGGAGA 926
778 GTCGCGCTCGTGGCTGCAATGCGCGCGCTCGTGGCGCGCAAGTACCGCGCGCA 837
927 GTTTGAAGAAAGTTGAATCTGTTTGAAGAAAGTTGAGGACGCTGAAGGCAAGTGTAT 986
838 GTTCGAGGAGCGGCTCAAGGCGGTGCTCAAGGAGGTGAGGAGCGCGACGCGCAAGGTAT 897
987 TCTCTTTATGATGAGATTCAATTTGGTTCTGTTGGTGTGGCAAACTGAAGGCTGATGGA 1046
898 CCTCTTATCAACGAGATACACCTCGTCTCGCGCGCGGGAGGAGCGAGGCGGTTCATGGA 957
1047 TGCAGCTAATCTGTTCAAGCCCATGTTAGCTPAGAGGCGAGCTTCGATGCTATTGGTGTAC 1106
958 CGCGGCAACCTCTTCAAGCCCATGCTCGCCAGGGGCGAGCTCGGTTGCTTCGCGCCAC 1017
1107 AACGCTTGAAGAAATACAGAAATATGTTGAGAAAGATGCTGCCCTTTGAGAGGAGTTCCA 1166
1018 CACGCTGAGGAGTACCAAGTACGTCGAGAAAGACGCGCGCTTCGAGCGACGCTTTCCA 1077
1167 ACAAGTCTATGTCGAGGACCAAGTGTGCTGACCACTTAGTATCTTATAGAGACTCAA 1226
1078 GCAGGTGACGTGCGCGGAGCGCGCTGCGCGACCACTACAGCATCTCGGAGGCTCAA 1137
1227 GGAGAGTATGAGGAGACATCATGTTGCTGCGAATCCAGACAGAGCTCTTATTAATGCTGC 1286

Db 1138 GGAGAGTACGAGGGGACCAACGCGGTGCGCATCCAGGACCGGCCATCGTCTGTCGCGC 1197
QY 1287 TCAGCTGTCTGCTGCTTACATAAATCTGGTCGGCATTTTACCGATAAGCAATTTGATTGCT 1346
Db 1198 GCAGCTCTCGGCAAGTATCATCTGGCGCGCATCTGCGGACAGGCCATCGACTGCT 1257
QY 1347 TGATGAGGCTTGTGCGAATGTGAGAGTCCAGCTTGATAGTCAACTGAGAGATTTGATAA 1406
Db 1258 GCACGAGGCGTCCGCAACGCTGAGGCTGAGTGGAGTGCACGCGCTGGAGAGGAGAAAGA 1317
QY 1407 CTTTGAAGAGAGAGTGCAGCTGGAAATTTGAATTTCAAGCTTGGAAAGGAGAAAGA 1466
Db 1318 TCTGAGCGAGAGCGATTCAGCTGGAGTGGAGTGCACGCGCTGGAGAGGAGAAAGA 1377
QY 1467 TAAAGCCAGCAAGCTCGACTTATAGAGTGGAGAGAGCTTGATGACCTGAGAGACAA 1526
Db 1378 CAAAGCCAGCAAGCGCGCTGGTGGAGTGGAGAGAGCTTGGACGACTTGGAGGACAA 1437
QY 1527 GCTTCAGCTCTCAAGATGAATAAGAAAGAGAGAAAGAGATTTGATGAGATTCGAAG 1586
Db 1438 GCTTCAGCGCTGACCATGAAGTACAGGAAGAGAGAGAGCGGATCGACGAGATCCGAA 1497
QY 1587 GCTTAAACAGAAAGAGAGAGCTCATGTTTTTTTTCAGGAGGCGAGAACGAAGATAGA 1646
Db 1498 GCTGAAGCAGCGCGCGGAGAGCTCGAGTTTCAAGTACGCGCGCTCCAGGAGATCGACGCGCATGCA 1557
QY 1647 CTTTCAGAGAGCTGCTGATCTTAAGATATGGCGCAATTCAGAAAGTGGAAATCTGCAATTCG 1706
Db 1558 CTTGCGCAGCTGCGCGACCTCAAGTACGCGCGCTCCAGGAGATCGACGCGCATTCG 1617
QY 1707 CAACTTGAAGAACTTCTTCTGAAGAGAAATGTGATGCTCACAGAAACGTTGGGCTGA 1766
Db 1618 GAAGCTGGAGGG ---CGAGACCGCGCGAGAACTGATGCTCACCGAGACCGTGGCGCCGA 1674
QY 1767 ACACATTTGCTGAGGTTGTGAGCGCTTGACAGGAGTTCAGTGAAGAGTTCGCGCAAAA 1826
Db 1675 GCAGATCGCGAGGTTGTGAGCGGTTGACCGGCTCCGCTGACCCGCTCGCGGACAA 1734
QY 1827 TGAGAGGAGAGGTTGATTTGCTTCTGATAGATGATGATTAAGCGGCTTGGGAGAGAA 1886
Db 1735 CGACAAGGAGCGCTGCTGCGCATGCGCGACAGGCTGCACAGAGGTTGCTCGGCGAC 1794
QY 1887 TCAGCGGTAAATGACAGTTTCTGAGGCAATTTCAAGTCAAGGCGAGACTTTGTTAGGCG 1946
Db 1795 GGAGGCGGTGAACGCGCTGCGCGAGCGGTTGCTGAGGTTCCAGGCGCGGCTTGGCGCGCC 1854
QY 1947 ACACAGAGCACTGCTCATTTCTTCTGAGCAACTGCTGTTGCAAACTGAGCT 2006
Db 1855 GCAGAGCCCAACGCGTCTTCTTCTGCGCGCGACCGGCTGGGCAAGAGCT 1914
QY 2007 CGCCAAAGGCTTCTGCTGAGCAGCTGTTGATGATGAAACCTTTAGTTTCGATGATAT 2066
Db 1915 CGCCAAAGGCTTCTGCGGAGCAGCTGTCGACGAGAACTCTGCTCGTCCGATCGACAT 1974
QY 2067 CTCGNAATATGGAACAACTCTGCTCTGCTCATTTGGGCGACCCAGCGGTATGT 2126
Db 1975 GTCCGAGTACATGGAAGCAGCTCCGCTCGCTGCTCATTCGCGCGCGCCAGCGGTATGT 2034
QY 2127 TGTCTACAGAGGAGTGGACAACTAACTGAGGCTGTGAGGAGGCGACCTTATTGTTGAT 2186
Db 2035 CGGACACAGAGAGGCGCGGCGCTGACGAGCAGGTGAGGAGGCGCGTACAGCGTAT 2094
QY 2187 ACTTTTGAATGAGTGGAGAGGCTCATGTTGCTGTCTTCAACACTCTGCTCCAGTTT 2246
Db 2095 CTTGTTGAGACGAGTGGAGAGGCGCAGTGGCGGTGTTCAACACGCTCTCTCCAGGTCT 2154
QY 2247 GATGATGCTCAATTTGACAGCGGCAAGGCGAGGCTCGATTTTCAGAACTCGGTGAT 2306
Db 2155 GGACGAGCGCGGCTGACCGAGCGGCAAGGCGGCTGAGCTTCGGAACACCGTAT 2214
QY 2307 AATCATGACATCAAACTTTGGTGTGTAACACTCTCTCTGAGGCTTAACTGGAAAGTAA 2366
Db 2215 CATCATGACCTCAAACTCTGCGCGCGCGACACTCTCTCTCGCGGATGTTGGGCAAGACTC 2274

2367 AATGAAGTGGCCCGGAGCTGTGTGATCGCGAGGTGAGGAAACACTTCAGACCAAGACT 2426
|||||
2275 CATGAAGGTGCGCCCGAGCTCTGTATGACGAGGTGAGGAGGCACTTCAGGCGGAGCT 2334
|||||
2427 CTTGAACAGGCTTGACGAGATGTGTGTGTCGACCCCTTTCACATGACCCAGTTCAGGAA 2486
|||||
2335 GCTGAACCGCTGACGAGATGTGTGTGTCGACCCCTTTCACATGACCCAGTTCAGGAA 2394
|||||
2487 AGTAGTCGGCTTCAAAAGAGAGCTGTCTGCTCGGCTTGTGCTGCTGCTGCTGCTGCTGCT 2546
|||||
2395 GGTGGCCCGCTGCAGATGAAGAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2454
|||||
2547 GCAGTCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2606
|||||
2455 GCGCGCTACCGGCTGAGCGGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2514
|||||
2607 TGCTAGGCTTATAAGGAGATGATGAGAGAGAGTGTGTGAGAGAGTGTGTGAGAGAGTGTGT 2666
|||||
2515 CGCTCGCCGATCAGGAGTGTGTGAGAGAGTGTGTGAGAGAGTGTGTGAGAGAGTGTGT 2574
|||||
2667 TGTGCTGAGGAAATCGATGAAGAACTTCACTGCTTTACATGATGAGGCG---CTGGTGA 2723
|||||
2575 GATCCAGGAGGAGATCGACGAGATTGCACTGTGTACATCGACGCGCGGCAAGGACGA 2634
|||||
2724 TCTTGTGTACCGGTAGA---AAGTGGAGTCTAGTGGACGCTTCAACAGGCAAGAGTC 2780
|||||
2635 GCTGGCTTACCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAG 2694
|||||
2781 AGATGTGCTGATTCTATTGCTTAACGGGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCT 2828
|||||
2695 GCATCTCTCATCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2754
|||||
2829 TCAGCGGTGAAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 2873
|||||
2755 CAAGCGGTGAAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 2799
|||||

ULT 14

HRIV39

US

INITIATION

SESSION

AL161539

AL161539.2

GI:7268147

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 19796)

EU Arabidopsis sequencing project.

Direct Submission

Submitted (10-MAR-2000)

MIPS, at the Max-Planck-Institut fuer

Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:

lenckemips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project

Coordinator: John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,

E-mail: michael.bevan@bbsrc.ac.uk

Information on performance of analysis and a more detailed

annotation of this entry and other sequences of chromosomes 3, 4

and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>

this fragment has an overlap with ATCHRIV40 at the 3' end and an

overlap with ATCHRIV38 at the 5' end.

Location/Qualifiers

1..197976

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/variety="Columbia"

/db_xref="taxon:3702"

/chromosome="4"

5668..7235

/gene="AT4g14420"

gene

gene

CDS

complement(join(5668..5772,5865..5963,6051..6140,
6238..6352,7168..7235))

/gene="AT4g14420"

complement(join(5668..5772,5865..5963,6051..6140,
6238..6352,7168..7235))

/gene="AT4g14420"

/note="strong similarity to Nicotiana tabacum ORF able to
induce HR-like lesions", PATX:G1762945
contains EST gb:AI999161.1"

/codon_start=1

/product="elicitor like protein"

/protein_id="CAB78484.1"

/db_xref="GI:7268148"

/translation="MSLASFLGRALFVSVLLSAWQFNDFGEDGGRSAKSLKRFNA
FVNVHTTHTQQQLPPVDMKILVAALAKIGLLFVFGSSLCAYLLILHQAATPIL
YDFNVYDVRKRFQSLKFTQSLALLGLLPIFMKNSRKHGRQLRKKAPKAKN"

complement(5668..5772)

/gene="AT4g14420"

/number=1

complement(5773..5864)

/gene="AT4g14420"

/number=1

complement(5865..5963)

/gene="AT4g14420"

/number=2

complement(5964..6050)

/gene="AT4g14420"

/number=2

complement(6051..6140)

/gene="AT4g14420"

/number=3

complement(6141..6237)

/gene="AT4g14420"

/number=3

complement(6238..6352)

/gene="AT4g14420"

/number=4

complement(6353..7167)

/gene="AT4g14420"

/number=4

complement(7168..7235)

/gene="AT4g14420"

/number=5

8407..9129

gene

gene

CDS

/note="similarity to carnitine racemase, Escherichia coli
contains EST gb:H36845, AI992737.1, T76583"

/codon_start=1

/product="carnitine racemase like protein"

/protein_id="CAB78485.1"

/db_xref="GI:7268149"

/translation="MCTLEKRGDLFLTLTGDEHRRPDDTTATLSLLEQAKSQSTR
GSILITANGKFFSNGFDLAWTAGSKTGAANRLHOMVEKFPVVAALLDLPMTTIA
ALNGHAAAGLILALSHDVFVRKDRGVLYMSEVDIGLSMPDYFSALVRKVTGSAAR
RELLSGKKIRGEAVGLGIVDVAAYDSEGVVASVRLGERLAKKWSGEVYVASIRK
SLYPELCGLGLETRVPATPKL"

complement(8407..9129)

/gene="AT4g14430"

/number=1

10443..11159

/gene="AT4g14440"

complement(10443..11159)

/gene="AT4g14440"

complement(10443..11159)

/gene="AT4g14440"

/note="similarity to carnitine racemase, Escherichia coli
contains EST gb:AI994308.1, AA394352, H36791, AA395864,
T76583, H77091, T46265, AA597888, Z17630"

/codon_start=1

gene

gene

CDS

FEATURES

source

[illegible]

1176 TGTTGGGAGCCAAAGTGTGCTGTGACACCAATTAGTATCTTAGAGGCTCAAGGAGAAAGTA 1235
 114848 TGTAGCTGAGCCTAGTGTGCTGCTGATCTATCAGTATCTTCGAGGGCTTAAAGAAAAATA 114907
 1236 TGAGGACATCATGCTGTGCGAATCCAGGACAGAGCTCTTATAAATCTGCTAGCTGTC 1295
 114908 TGAAGGACATCATGCTGTGCTGATATCCAGATAGAGCTAGTCTTTCTGCTCAATTGTC 114967
 1296 TGCTCGTTACATAACT 1311
 114968 TGAACGTTTACATCACTGCTATGTTGCGCCATCTTTTATCTAACATGCTCTTGTAAATTA 115027
 1312 115028 TTATGATTTCTTTATGTTGCTTATTAGTGTCTGTGCTGTAGTCTGAGCTTACC 115087
 1326 GGATAAAGCAATTCATTTGTTGATGAGGCTTGTGCGAATCTGAGAGCTCAGCTTGATAG 1385
 115088 GGATAAGCAATTCATTTGTTGATGAGTCTTGTGCAATGTTCAAGCCAGCTTGATAT 115147
 1386 TCAACCTGAAGAGATTGATAACCTTTGAAGGAAGAGGATGAGCTGGAATTTGAACCTCA 1445
 115148 TCAACCAAGAAATTCATAGCTTTGAAGAAAGTAATGCAACTAGAGATAGAGATTCA 115207
 1446 CGCCTTGGAAAGGAGAG---GATTAAGCCAGCAAGAGCTCGAC 1486
 115208 TGCTCTCGAAAGGAGAGGACGCAAGAGCTAGCGAAGCTCGTCTTCCGAGGTAAATTA 115267
 1487 115268 CTTTCAAAACATCGTGTTTTACATATAGTACCTCACTCTTTTATTTGTTATTTCTGCAGG 115327
 1496 TGGGAAAGAGCTTGATGACCTGAGAGACAGCTTTCAGCCCTCAAGATCAAAATACAGAA 1555
 115328 TGAGAAAGAGCTTGATGATCTTGAGGAGCAAGCTTGAAGCTCTTACAAATAAATAAATA 115387
 1556 AGGAAAGAGAGAAATTCATGAGATTCGAGGCTTAAACAGAAAGAGAGAGCTCATGT 1615
 115388 AGGAAAGAGAGATCAATTAATGAACCTCGAGGCTTAAACAGAAAGAGAGAGCTCATGA 115447
 1616 TTCTTTTGGAGGAGGAGAGAGATATGACCTTGCAAGAGCTGCTGATCTAGATATG 1675
 115448 TTGCTCTACAGAGGCTGAGAGACATGACGTTTCTAAAGCTGCTTCTAAATATG 115507
 1676 GCGCAATTCAGAAAGTGAATCTGCAATTCGCCAACTTGAAGGAACTTCTTCTGAAGAGA 1735
 115508 GTGCAATTCAGAAAGTGAATCTGCGATAGCGAACT---AGAAAGAGTCTTAAAGATA 115564
 1736 ATGTGATGCTCACAGAAACGTTGGGCTGAAACATTCGCTGAGGTTGTGAGCCGTTGGA 1795
 115565 ACGTGTGCTTACAGAAACGTTGGGCGAGAGATATTTGCTGAGGTAGTGGCCGATGGA 115624
 1796 CAGGATTTCCAGTGCAGAGACTTGGCCAAATCAGAAAGAGAGGTTGATTTGCTTGTGCTG 1855
 115625 CTGGATACAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 115684
 1856 ATAGTTTGCATAAGCGGTTGTGGGACAGAAATCAAGCGTAAATGCAAGTTTCTGAGGCAA 1915
 115685 ATAAGTTGCAGAGAGAGAGTTGTTGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 115744
 1916 TTCTAAGTCAAGGCGAGAGCTTGTTAGGAGCAACAGCCCACTGAGATCATTTTATTC 1975
 115745 TACTAAGGTCAAGGTTGGACTAGGAGAGGCAACAAACCGCTGCTTCTCTCTCC 115804
 1976 TTGACCAACTGTTGGCAAACTGAGCTCGCAAGCTCTTGTGAGCAGCTGTTG 2035
 115805 TCGGTCCAACTGTTGGTAAACCGAGCTGAGCTGAGGCTGAGCAACTTTG 115864
 2036 ATGATGAAACCTCTTAGTTGCGGATGATATGTCGGAATATATGGAACCAACACTCTGTCT 2095
 115865 ACAGTGAAACCTTACTCGTTAGGCTTGATATGTCGAGATATATGACAAATTTCTTGTTA 115924

QY 2096 CTCGCTCATTTGGGCGACCAACAGGGTATG 2125
 Db 115925 ATAAACTTATAGGGCGACCACTGGATAGCTGCTTAAACTCAAGCCACTTCTTTATGTGT 115984
 QY 2126 2126
 Db 115985 TTTTGTGATTAATGTTATATGGAACCTCACATGATACTACTTTTGTTTTCTTGTAGTACAT 116044
 QY 2127 TGTCTACAGGAGAGGTGGACAACCTAACTGAGGCTGTGAGGAGCGACCTTATTTGTGTCTAT 2186
 Db 116045 TGTCTATGAGAGAGGTGGACAACCTAACTGAGGCTGTGAGGAGCGACCTTACTGTGTGT 116104
 QY 2187 ACTCTTTGATGAAGTGGAGAGGCTCATGTTGCTGCTCTTCAACACTCTGCTCCAAAGTTT 2246
 Db 116105 ACTATTTGATGAAGTTGAAAGACCCATGTTACTGTTTCAACAGCTACTTCAAGTGTCT 116164
 QY 2247 GGATGATGCTCAATTCAGACAGCGGCAAGGAGGAGGACAGTCTGATTTTCAGAACTCGGTGAT 2306
 Db 116165 AGAAGACGGGAGATTAAACCGATTAGCATGGGAGGACAGTCGATTTTAAAGAACACAGTGTAT 116224
 QY 2307 AATCATGACATCAAAACCTTGGTGTGTAACACCTCTTCAGGGCTAACTGGGAAAGTAAC 2366
 Db 116225 CATATGACTTCTTAACCTTGGTGTGAGACCACTTGTTCAGGGCTTACAGGTGAATGAC 116284
 QY 2367 AATGAAAGTGGCCCGGACTGTGTGATGGGAGGTGAGGAAACACTTTCAGACCAAGCT 2426
 Db 116285 AATGCAAGTGGCTCGAGACCAACGCAATGAAGATGCGAATAAACAATTTTCAGCCAGAGCT 116344
 QY 2427 CTTGACAGGCTTGACGAGATGTTGTTTGCAGCCCTTTTCATGATGACCAAGTTCAGGAA 2486
 Db 116345 GTTAAACAGATTGGATGATGATGTTTATGTTTTCATCTTTGTCACGAGCATTTGGCAAA 116404
 QY 2487 AGTAGTCTGGCTTCAAAATGAAGAGCTTGTGTCGGCTTGTCTGAAGAGGAGTTGCTTT 2546
 Db 116405 AATTGTTCACTTCAAGTGAGAAATGTTGCT 116437
 QY 2547 GGCAGTCTACTGATGCTGTTTGGACTATATCTTGGCAGAGAGTTATGACCGGTGTATGG 2606
 Db 116438 -----TGCTTGAAGTTTACATTTTATAATATGATGATGTTTATTTAGTGTATGG 116492
 QY 2607 TGCT--AGGCTTATAGGAGATGATGAGAGAGAGGTTGTTGACAGAACTGTCAAAGT 2664
 Db 116493 TGCTAGAGGCTTATTAAGATGCTGAGAGAGAGAGGTTGTTGACAGATATATCAATGATG 116552
 QY 2665 GTTGTGCTGAGGAAATCGATCAAACTCCACTGTTTACATAGATGACGGCGCTGTTGAT 2724
 Db 116553 ATTGTGCGAGAGAAATCGGTGACGATTTCCATTTGTCATGATGTTAAACGAGGCTAAA 116612
 QY 2725 CTTGTGTACCGGTTAGAAAGTGGAGGCTTAGTGACGCTTCAACAGGCAAGAGTCAAGT 2784
 Db 116613 ACTGATCTTGTCTACCGAATAGATAAAATGTTTTTGTGAAGATAGAGCAACATTAGAT 116672
 QY 2785 GTGCTGATTCATTTGCTTAAACGGCCCAAGAGAGTGTATG 2824
 Db 116673 GTTGTGATTCATTCTGGAATTAAGAGAGAGAGAGCAATG 116712

RESULT 15
 Locus ATFCAL
 DEFINITION Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
 No. 1.
 ACCESSION Z97336
 VERSION Z97336.1
 KEYWORDS GI:2244788
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
 1 Bevan, M., Stiekema, W., Murphy, G., Wambutt, R., Pohl, T., Terry, N.,
 Kreis, M., Kavanagh, T., Entian, K.D., Rieger, M., James, R.,

Puigdomenech, P., Hatzopoulos, P., Obermaier, B., Duesterhoft, A.,
 Jones, J., Paime, K., Ansgore, W., Delseny, M., Bancroft, I.,
 Meyers, H. W., Schueller, C. and Chalwatzis, N.
 Unpublished
 2 (bases 1 to 206606)
 EU Arabidopsis sequencing, project.
 Direct Submission
 Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitze 18a, D-82152 Martinsried, FRG, E-mail:
 schueller@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de, Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@bbsrc.ac.uk
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
 this fragment has an overlap with AtFCA0 at the 5' end and an
 overlap with AtFCA2 at the 3' end.

FEATURES

source	Location/Qualifiers	
	1..206606	
	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/variety="Columbia"	
	/db_xref="taxon:3702"	
	/chromosome="4"	
source	1..26121	
	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="BAC TAMU8H22"	
source	12367..32771	
	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="cosmid YUP16B9.IB1N11"	
source	32771..36017	
	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="cosmid YUP16B9.IB1D3"	
source	33868..135899	
	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="BAC TAMU8P4"	
source	134292..149379	
	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="cosmid YUP16B9.IB1F18"	
source	146795..164143	
	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="cosmid YUP16B9.IB1K18"	
source	162663..193937	
	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="cosmid YUP16B9.IB1N23/YUP16B9.IB1O24"	
source	183320..206606	
	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:3702"	
	/clone="BAC TAMU2D4"	
gene	685..1806	
	/gene="dl3200c"	
gene	complement(join(685..744,829..903,1053..1232,1804..1806))	
	/gene="dl3200c"	
CDS	complement(join(685..744,829..903,1053..1232,1804..1806))	
	/gene="dl3200c"	
	/note="strong similarity to ribosomal protein L41, Candida maltosa"	
	/codon_start=1	
	/product="ribosomal protein"	
	/protein_id="CAB10211.1"	
	/db_xref="GI:2244789"	
	/translation="MVNPKTKNTYCKNCKKHTLHKVTQYKKGKDSLAAGKRRYD RKQSGYGGGTQKPVFHKAKTKKIVLRQCQSKFSPRIKCKHFFETGGKGGKGT SLF"	
exon	complement(685..744)	
	/gene="dl3200c"	
intron	number=1	
	complement(745..828)	
	/gene="dl3200c"	
exon	number=1	
	complement(829..903)	
	/gene="dl3200c"	
intron	number=2	
	complement(904..1052)	
	/gene="dl3200c"	
exon	number=2	
	complement(1053..1232)	
	/gene="dl3200c"	
intron	number=3	
	complement(1233..1803)	
	/gene="dl3200c"	
exon	number=3	
	complement(1804..1806)	
	/gene="dl3200c"	
gene	2229..7384	
	/gene="dl3205w"	
CDS	join(2229..2837,2914..3069,3161..3268,3358..3468, 3569..4957,5000..5212,6340..6428,6469..6560,6706..6749, 7316..7384)	
	/gene="dl3205w"	
	/note="similarity to kinesin osm-3, Caenorhabditis elegans contains EST gb:496806"	
	/codon_start=1	
	/product="kinesin like protein"	
	/protein_id="CAB10212.1"	
	/db_xref="GI:2244790"	
	/db_xref="SPTREMBL:O32291"	
	/translation="WAPTPSSRSNQYTLIRTPQKRLNFKHTNPNGSKDPSP PEHVEVIGIRIDYDPRKEKSPILQVNDQTVRVADGVYDFLDGVFSQEG BEFYKPIERIKGVKGNKCTIMMYGPTGAKSHTFGCGKEGIVYVSRDLILGDS DQGVTFVQTVLVNBEIYDLSTNSNNGICMPKGAATKVELEVKKAKNAF ISSTAGKISKEIVKVEKRIKSTLCHERSRSCHILIDVPTVGGRLMLVDAGSE NIDAGQGFEMKQAKINQNIALKEVRSIANGSDSHVPRDSKLTMLLODSFDD KSKMLICASPDPKEMHKLCTLEYKAKCIVRGSHIPNKKYGGESASAVILGS RIAAMDEFITIKOSEKKOKERNEAKOLKKEEVEVALRSLLTORACATNEEIK EKVNERTQLKSELDKLEECRMAEEFVEMRMEERIIVQQBELEMMRRLEIE VEFRRNGSVDETSGFARLSLYSDDDPGMVKSMIDMGDPEPVQVWGAHVHQS NTFISGNTLLOKPSNNLTOMYPDRCLSTVPEEVEVEEKVIVEDKSIICLIT PMSINSEGLKGCNFGADKESASRLRLQNTFTLQGNQRELSQHSQGOEQDNI ASPDKDNQPFSTINKAALAVEAKENNISVDQRENGQLDIYVKNETAADNPKLIT LTRVTDATLADRLKLIIEYIGSDNQAFELKIGVINLQQAQAFHFILPECGAQV AKEKSTVQATSLRPLCNGHAYLALRPGKSHQSLQSPASPLPLNPIENMEVTPISK VTFNHQILKTLERIGADYINLSLHFLILFPRGESHQHRLMSPHSHQTSPLSS FPLLSRADBPGLVDITPLFEVVLGFTLVLELAVVNDRIAG"	
exon	2229..2837	
	/gene="dl3205w"	
intron	number=1	
	2838..2913	
	/gene="dl3205w"	
	number=1	
exon	2914..3069	
	/gene="dl3205w"	
intron	number=2	
	3070..3160	
	/gene="dl3205w"	
	number=2	
exon	3161..3268	
	/gene="dl3205w"	

[illegible]

Db 171177 GTTGTGATTCAATTCAGGGNAATAGAGAGGAAGCAATG 171216

Search completed: February 12, 2004, 22:27:41
Job time : 11053 secs

```

170129 CTGGATACCACTGACTAGACTTGCACAGACGAGAGAGAGGCTAATTTCTCTGCTG 170188
1856 ATAGGTTGCATAAGCGGTTGTGGACAGATCAAGCGTAATAGCAGTTCTGAGCAA 1915
170189 ATAGTTGCACGAGAGAGTTGTGGACAGAGAGGCTGTAAGAGCTGTGGCTGCTCAA 170248
1916 TTCTAAGGTCAAGGCGAGAGCTTGTAGGGCACACAGCCAACTGGATCAATTTCTATTCC 1975
170249 TACTAAGGTCGAGGGTTGGACTAGGAGGCCACACACCGTCTGGTTCTCTCTTCC 170308
1776 TTGACCAACTGGTGTGGCAAACTAGCTCGCAAGGCTCTTGTGAGAGCTGTTTG 2035
170309 TCGGTCCAACCTGGTGTGGTAAACCGAGCTGGCTAAGGCTCTGGCTGAGCAACTCTTTG 170368
2036 ATGATGAAACCTCTTACTGCGATTGATATGTCGGAATATATGAAACAACTCTCTCT 2095
170369 ACAGTGAACCTTACTGTTAGGCTTGTATATGTCGAGTATATGACAAATTTCTCTGTTA 170428
2096 CTCGCTCATTTGGGCGACACACAGGGTATG----- 2125
170429 ATAACTTTATAGGGGACCACTTGGATAGCTGTAAACTCAAGCCACTTCTTTATGTGT 170488
2126 -----T 2126
170489 TTTTGTATTAATGGTATTAATGGAACCTCACATGATACTTCTTTTCTTAGGTACAT 170548
2127 TGGTCACGAGGAGGTGACAACTAACTGAGGCTGTGAGAGGCGACTTATTTGTGTCTAT 2186
170549 TGGTCATGAGAGGTTGACAACTAACTGAACTGTGAGAGGAGCACTTACTGTGTGT 170608
2187 ACTCTTTGATGAAGTGGAGAGGCTCATGTTGCTGTCTTCAACACTCTGTCTCCAAGTTT 2246
170609 ACTATTTGATGAAGTTGAAAGACCCATGTAACTGTATTTCAACAGCTACTTCAAGTGCT 170668
2247 GGTGATGCTGCTGATGACAGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2306
170669 AGAGACGGGAGATTAAACCGATTAGCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 170728
2307 AATCATGACATCAAACTTGTGTGTGTAACACTCTTGTGAGGGCTAATCGGAAAGTAAC 2366
170729 CATATGACTTCTAACCTTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 170788
2367 AATGAACTGCGCGGAGCTGTGTGATCGGGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 2426
170789 AATGCAAGTGGCTCGAGACAAAGCAATGAAAGATGCGAAAAAATTTCAAGCCAGAGCT 170848
2427 CTGAAACAGGCTTGACGAGATTGTGGTGTGCGACCCCTTTCACATGACACGAGTTGAGGAA 2486
170849 GTTAAACAGATTGGATGAGATGATTATTTTATATATTTATATGATGATTTGTTAGGTGATGG 170908
2487 AGTAGCTCGGCTTCAAATGAAAGAGGTTGCTGTGCGGCTTGTGAAAGAGGAGTTGCTTT 2546
170909 AATGTTCAACTCCAAAGTGAAGATGTTGTCT----- 170941
2547 GGCAGTCACTGATGCTGCTTTGGACTATATCTTGGCAGAGAGTTATGACCCGGTGTATGG 2606
170942 -----TGCTTGGAGTTTACATTTTATATATTTATATGATGATTTGTTAGGTGATGG 170996
2607 TGCT--AGGCTTAAAGAGATGGATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2664
170997 TGCTAGAGGCTTATGAGAGATGGCTAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 171056
2665 GTTGTGCGTGAGGAAATCGATGAAACTCCACTGTTTACATAGATGCGAGCGCTGGTGAT 2724
171057 ATTTGTGCGAGAGAAATCGGTGACGATTTCCATTTGTGCTATTTGATGTTAAACGAGCTAAA 171116
2725 CTTGTGTACCGGGTGAAGAGTGGAGGCTGTAGTGGAGGCTTCAACAGGCAAGAGTCAGAT 2784
171117 ACTGATCTTTGCTACCGGAATAGATAAAATGTTTTGTGAAGATAGAGCAAAACATTAGAT 171176
2785 GTGCTGATTCTATTTGCTAAACGGGCAAGAGAGGAGTATG 2824

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - nucleic search, using frame_plus_p2n model

on: February 13, 2004, 01:34:21 ; Search time 535 Seconds
(without alignments)
4596.613 Million cell updates/sec

le: US-09-812-350-17
fect score: 4550
uence: 1 MNPEKFTHTNETIATAHEL.....VKXWRIEIEDDDNEEMIED 911

ring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

rchd: 2552756 seqs, 1349719017 residues

al number of hits satisfying chosen parameters: 5105512

imum DB seq length: 0

imum DB seq length: 2000000000

-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

mand line parameters:

DEL=frame+ p2n.model -DEV=xlp
cg2 1/USPTO spool_P/US09812350/runat 10022004_101514_13343/app_query.fasta_1.1095
N Geneseq 19Jun03 -QPMFASTAP -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
PEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
ST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
BE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
CGN 1 1 396 @runat 10022004_101514_13343 -NCPU=6 -ICPU=3
MMAP -LARGEOQUERY -NEG SCORES=0 -WAIT -DSELOCK=100 -LONGLOG
TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
PEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

abase : N Geneseq 19Jun03:
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4550	100.0	3105	22	AAI66065	Arabidopsis thalia
2	4545	99.9	2736	24	ABZ12430	Arabidopsis thalia
3	4545	99.9	2736	25	ABZ42033	Arabidopsis thalia
4	4388	96.4	6376	22	AAI66066	Arabidopsis thalia
5	4003	88.0	3049	22	AAI66067	Glycine max heat s
6	3896	85.6	3052	22	AAI66068	Nicotiana tobacum
7	3883.5	85.4	3084	22	AAI66070	Zea mays heat shoc
8	3822.5	84.0	3058	22	AAI66069	Triticum aestivum
9	3790	83.3	2821	22	AAI66075	Triticum aestivum
10	3657	80.4	3942	22	AAI66073	Phaseolus lunatus
11	3657	80.4	3942	22	AAI66074	Triticum aestivum
12	2502.5	55.0	2045	22	AAI66071	Zea mays 101 kDa h
13	2502.5	55.0	2045	22	AAI66072	Zea mays heat shoc
14	2298	50.5	7726	22	AAI66083	Leishmania donovan
15	2289	50.3	5997	22	AAI66084	Leishmania major h
16	2288	50.3	2619	24	ABS63272	High level promote
17	2245.5	49.4	43980	24	ABQ67192	Listeria innocua c
18	2245.5	49.4	3011208	24	ABQ69245	Listeria innocua D
19	2236.5	49.2	2944528	24	ABA031041	Listeria monocytog
20	2224.5	48.9	2800	22	AAI66064	Trypanosoma brucei
21	2224	48.9	2776	22	AAI66057	Synechococcus CLPB
22	2224	48.9	2776	22	AAI66076	Synechococcus sp h
23	2204	48.4	2586	23	AAI66076	Salmonella typhi D
24	2194	48.2	2574	20	AAI66071	Staphylococcus aur
25	2193	48.2	2574	23	AAI66071	E. coli DNA for ce
26	2193	48.2	2574	24	AAI66071	E. coli clpB gene.
27	2185	48.0	34980	24	ABQ81847	Bifidobacterium lo
28	2185	48.0	34980	24	ABQ81848	Bifidobacterium lo
29	2162.5	47.5	2571	23	AAI66076	Haemophilus influe
30	2162.5	47.5	1830121	17	AAT42063	Mycobacterium tube
31	2158.5	47.4	2544	24	ABT05624	Mycobacterium tube
32	2158.5	47.4	4403765	22	AAI99683	Mycobacterium tube
33	2158.5	47.4	4411529	22	AAI99682	Mycobacterium tube
34	2150.5	47.3	3464	22	AAI66062	Corynebacterium gl
35	2150.5	47.3	349980	22	AAH68533	C glutamicum codin
36	2145	47.1	2556	22	AAH68464	C glutamicum codin
37	2141.5	47.1	2607	23	AAI66065	Staphylococcus aur
38	2141.5	47.1	2649	23	AAI66065	Staphylococcus aur
39	2140.5	47.0	23474	23	AAI66065	Propionibacterium
40	2136.5	47.0	6491	20	AAI66065	Enterococcus faeca
41	2136.5	47.0	6491	24	ABS98888	Enterococcus faeca
42	2133	46.9	2565	23	AAI66065	Pseudomonas aerugi
43	2133	46.9	2580	20	AAI66065	Nucleotide sequenc
44	2133	46.9	42235	20	AAI66065	Nucleotide sequenc
45	2111.5	46.4	2571	24	ABQ90177	M. capsulatus gene

ALIGNMENTS

RESULT 1

AAI66065
ID AAI66065 standard; DNA; 3105 BP.

AC AAI66065;

14-JAN-2002 (first entry)

Arabidopsis thaliana heat shock protein 101 gene (GenBank: U13949).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
Arabidopsis thaliana; ds.

Arabidopsis thaliana.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

P-PSDB; AAM51659.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAM51651-AAM51671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 3105 BP; 891 A; 541 C; 852 G; 821 T; 0 other;

Alignment Scores:

Seq. No.:	0	Length:	3105
ore:	4550.00	Matches:	911
Percent Similarity:	100.00%	Conservative:	0
Local Similarity:	100.00%	Mismatches:	0
Very Match:	100.00%	Indels:	0
:	22	Gaps:	0

-09-812-350-17 (1-911) x AAI66065 (1-3105)

1	MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu	20
163	ATGAATCCAGAGAAATTCACACACAGACAAACAGACAAATTCGTACAGCTCATGAGCTA	222
21	AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer	40
223	GCCTGGAATCAGGACATGCTCAATTCACCTCCTTTCGATTTAGCTGGTCTTGTCTCT	282
41	AspProThrGlyPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln	60
283	GATCCCAACCGGTATATTTCTCAAGCAATCTCTAGTCCGGTGGCGAGAACGCGAGCTCAA	342
61	SerAlaGluArgValIleAsnGlnAlaLeuLysAlaLeuProSerGlnSerProProPro	80
343	CTCTGCTGAAGAGTGATCAATCAAGCCCTTGAAGAGCTTCCTTCACAACTCTCTCCACCT	402
81	AspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGln	100
403	GATGATATTCAGCGAGTCTTAGTCTTTATTAAGTCAATTCGTGCTCAAGCTCTCAG	462
101	LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuAsp	120
463	AAGTCACGAGTGATCTACTATTTGGCTGTGTGACAGTTCGATTTAGGCTCTCTTGAAGAT	522
121	SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu	140

Db	523	TCTCAATCAGGATTTTGTGAACGAAGTCGTGTAGCGAGCGGAGGTAAAGTCTGAG	582
QY	141	ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn	160
Db	583	GTGGAAGCTTCGTGGGAAAGAGCGGAAGAGTGTAGAGTGTCTTCAGGGGACACAAAT	642
QY	161	PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro	180
Db	643	TTTCAAGCTTTAAGACTTATGGAAGAGATTGGTTGAGCAAGCAGGAGGAGCTTGATCT	702
QY	181	ValIleGlyArgAspGluGluIleArgValValArgIleLeuSerArgThrLys	200
Db	703	GTGATTGCTGCTGATGAGGAGATTAGAAGAGTCTGTAGAGATTCTTTCAGGAGAACGAAG	762
QY	201	AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu	220
Db	763	AACAATCTGTGCTTATTGGAGAGCCAGGAGTGTGTAACACAGCTGTGGTTGAAGGTTTA	822
QY	221	AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer	240
Db	823	GCACAAAGGATTGTGAAGAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAAATTCG	882
QY	241	LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluArgLeu	260
Db	883	TTGGACATGGGTGGCTTAGTGTGCTGCTAAATACCGAGGAGATTGGAAGAAAGTTG	942
QY	261	LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu	280
Db	943	AAATCTGTTTGAAGAGATTGAGAGAGCTGAGAGCAAGTGATTCCTTTATTTGATGAG	1002
QY	281	IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe	300
Db	1003	ATTCATTTGGTCTTGGTGTGCTGCAAAACCTGAAGGGTCGATGATGCAGCTAATCTCTTC	1062
QY	301	LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGluTyr	320
Db	1063	AAGCCCATGTTAGCTAGAGGCGCTTCGATGCAATTCGTGTCTACACGCTTGAAGAATAC	1122
QY	321	ArgLysTyrValGluLysAspAlaPheGluArgArgPheGlnGlnValTyrValAla	340
Db	1123	AGGAATATGTTGAGAAAGATGCTGCTTTGAGAGAGAGTTCACACAGTCTATGTTGCG	1182
QY	341	GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGly	360
Db	1183	GAGCCAGTGTGCTGACACCATTAGTATCTTAGAGAGCTCAAGAGAGAGTATGAGGA	1242
QY	361	HisHisGlyValArgIleGlnAspArgAlaIleLeuIleAsnAlaAlaGlnLeuSerAlaArg	380
Db	1243	CATCATGTTGTGCGAATCCAAAGCAGAGCTCTTATAATGCTGTCTCAGCTGTCTGCTCGT	1302
QY	381	TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla	400
Db	1303	TACATACTGTGTGCGCATTTACCGGATAAAGCAATTGATTTGTTGATGAGGCTTGTGCG	1362
QY	401	AsnValArgValGlnLeuAspSerGlnProGluGluIleAspLeuArgLysArg	420
Db	1363	AATGTGAGAGTCCAGCTTGTAGTAGTCAACCTGAAGAGATTGATTAACCTTGAAGAGAGG	1422
QY	421	MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla	440
Db	1423	ATGCAGCTGGAAATTTGAACCTTCACGCTTGGAAAGGAGAGGATAAAGCCAGCAAGCT	1482
QY	441	ArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThr	460
Db	1483	CGACTATAGAGGTGCGGAAGAGCTTGTGATGACCTGAGAGACAGCTTTCAGCTCTCACG	1542
QY	461	MetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArg	480
Db	1543	ATGAATACAGAAAGAGAGAGAGAGATTGATGAGATTGGAAGGCTTAAACAGAAAGA	1602
QY	481	GluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAlaArgAlaAla	500

1603 GAAGAGCTCATGTTTCTTTCTTTCAGAGGAGCAGAAAGATATGACCTTGGCAAGAGCTGCT 1662
 501 AspLeuArgTyrGlyAlaIleGlnValGluSerAlaIleAlaGlnLeuGluGlyThr 520
 1663 GATCTAGATATGCGCAATTCAGAAAGTGAATCTGCAATTTGCCAACTTTGAGGAAC 1722
 521 SerSerGluGluAenValMetLeuThrGluAenValGlyProGluHisIleAlaGluVal 540
 1723 TCTTCTGAAGAGAATGTGATGCTCACAGAAACGTTGGGCTGGAACATTTGCTGAGGTT 1782
 541 ValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAenGluLysGluArgLeu 560
 1783 GTGAGCCGTTGGACAGGATTCCTAGTACGAGACTTGGCCAAATGAGAGGAGGTTG 1842
 561 IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAenGlnAlaValAspAla 580
 1843 ATTGGTCTTGCTGATAGTTGCATAAGCGGTTTGGGACAGATCAAGCGGTAAATGCA 1902
 581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
 1903 GTTTCCTGAGGCAATTCCTAAGGTCAGGCGAGGACTTGTGTAGGCGACACAGCCAACTGGA 1962
 601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
 1963 TCATCTTATTCCTTGACCAACTGGTGTGGCAAACTGAGCTCGCCAAAGGCTCTTGCT 2022
 621 GluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluTyrMetGlu 640
 2023 GAGCAGCTGTTTGTATGATGAACCTCTTAGTTCGGATTGATGTCGGATATATGGA 2082
 641 GlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHisGluGluGly 660
 2083 CAACACTCTGCTCTCGCTCATTTGGGCGCACACAGGCTATGTTGTCACAGGAGGT 2142
 661 GlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGluVal 680
 2143 GGACACTCACTAGGCTGTGAGGAGCGACTTATTGTGTACACTCTTTGTATGAAGTG 2202
 681 GluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGlyArgLeu 700
 2203 GAGAAGGCTCATGTTGCTGCTTCAACACTCTGCTCCAGTTTGGATGATGTCGATG 2262
 701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleMetThrSerAsn 720
 2263 ACAGCGGCAAGCAGCAGAGTCAATTCAGGAACCTCGGTGATAATCATGACATCAAC 2322
 721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740
 2323 CTGCTGTGAACACCTCTTCAGGGCTAACTGGAAAGTAACATGGAAAGTGCCCGG 2382
 741 AspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAsp 760
 2383 GACTGTGTGATGCGGAGGTGAGGAACACTTCAGACCAGAGCTCTTGAACAGGCTTGAC 2442
 761 GluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGln 780
 2443 GAGATTGGTGTTCGACCCCTTTCATCATGACAGTTGAGAAAGTAGTCGCTTCAA 2502
 781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800
 2503 ATGAAGACGTTGTGCTCCGGCTTGTCTGAAGAGAGTGTCTTTGGCAGTCACGTATGCT 2562
 801 AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820
 2563 GCTTTGGACTATATCTTTGGCAGAGAGTTATGACCGGCTGTGCTAGCCCTATAAGG 2622
 821 ArgTrpMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIle 840
 2623 AGATGGATGGAGAGAGGTGGTCAACAGACTGTCAAGAGATGTTGTGCGTGAAGAAATC 2682
 841 AspGluAenSerThrValTyrIleAspAlaGlyValAspLeuValTyrArgValGlu 860
 2683 GATGAAGAACTCCACTGTTTACATAGATGACGGGCGCTGGTGTCTTGTGTACCGGAGAA 2742

QY 861 SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla 880
 Db 2743 AGTGGAGGTCTAGTGGACCTTCAACAGCAAGAGTCAGATGCTGATTATTTGCT 2802
 QY 881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIle 900
 Db 2803 AACGGGCCAAAGAGAGTGTGATGCTCAGGCGGTGAAGAAGATCAGGATCGAGGAATA 2862
 QY 901 GluAspAspAsnGluGluMetIleGluAsp 911
 Db 2863 GAAGATGACGATTAATGAGGAATGATCGAGGAT 2895

RESULT 2

ABZ12430
 ID ABZ12430 standard; DNA; 2736 BP.

XX ABZ12430;

XX DT 21-JAN-2003 (first entry)

DE Arabidopsis thaliana stress regulated gene SEQ ID NO 235.

XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

OS Arabidopsis thaliana.

PN WO200216655-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26685.

PR 24-AUG-2000; 2000US-227866P.

PR 26-JAN-2001; 2001US-264847P.

XX 22-JUN-2001; 2001US-300111P.

PA (SCRI) SCRIPES RES INST.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Harper JF, Kreps J, Wang X, Zhu T;

XX WPI; 2002-304127/34.

DR Identifying a stress condition to which a plant cell has been exposed and producing plants with increased tolerance to these abiotic stresses

XX Claim 144; SEQ ID NO 235; 577pp + Sequence Listing; English.

CC The invention relates to identifying a stress condition to which a plant cell has been exposed, comprising:

CC (a) contacting nucleic acid representative of expressed polynucleotides in the plant cell with an array or probes representative of the plant cell genome; and

CC (b) detecting a profile of expressed polynucleotides in the plant cell characteristic of a stress response. The method is useful in the

CC production of transgenic plants, cells and seeds and in producing plants with increased tolerance to abiotic stress. The present sequence is that

CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used in methods of the invention.

CC Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office.

XX SQ Sequence 2736 BP; 784 A; 481 C; 785 G; 686 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 2736

Score: 4545.00 Matches: 910

Percent Similarity: 99.89% Conservative: 0

Best Local Similarity: 99.89% Mismatches: 1

Query Match: 99.89% Indels: 0

09-812-350-17 (1-911) x AB212430 (1-2736)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
1 ATGAATCCAGAGAAATTCACACACAGACAAACGAGCAAAATTCGTACAGCTCATGAGCTA 60
21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
61 GCTGTGATCGACAGACATGCTCAATTCATCTTCCTTCGATTTAGCTGGCTTTGATCTCT 120
41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGluAsnAlaAlaGln 60
121 GATCCCAACCGGTATATTCCTCAAGCAATCTCTAGTCCGGTGGCGAGAACGAGCTCAA 180
61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
181 TCTGCTGAAAGAGTGATCAATCAAGCGCTTGAAGAGCTTCCTTCAATCTCTCCACCT 240
81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaGln 100
241 GATGATATTCAGCGAGTTCATGCTTATTAAGTCAATTCGTCGCTCAAGCTGCTCAG 300
101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuAsp 120
301 AAGTCACGAGGTGATATCTATTTGGCTGTTGACCACTGATTATGGGTCTCTTTGAAGAT 360
121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
361 TCTCAATCAGGAGTTTGTGAAGAGTGGTGTAGCGACGGCGAGGGTAAAGTCTGAG 420
141 ValGluLysLeuArgGlyLysGluGlyLysValLysValGluSerAlaSerGlyAspThrAsn 160
421 GTTGAGAGCTTCGTGGGAAAGAGGAGGAGAGTTCAGAGTCTTCAGGGGACACAAAT 480
161 PheGlnAlaLeuLysThrThrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
481 TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCGAGGAAAGCTTATCT 540
181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgThrLys 200
541 GTGATTTGGTCTGATCAGGAGATTAGAAGAGTCTGAGGATTTCTTCGAGGAGAGCGAAG 600
201 AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220
601 AACAACTCTGTGCTTATTCGAGAGCCAGGAGTTGGTAAACAGCTGTGTTGAAGTTA 660
221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
661 GCACAAAGGATTTGTGAAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTGC 720
241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysThrArgGlyGluPheGluGluArgLeu 260
721 TTGGACATGGTTCGCTTAGTTCGTGCTGCTTAATACCGAGGAGAGTTTGAAGAAAGTTG 780
261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
781 AAATCTGTTTGAAGAGTTCGAGGAGCTGAGGCAAGTCAAGTCAATCTCTTTATTATGAG 840
281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300
841 ATTCATTTGGTCTTCGTGCTGCAAAATCGAAGGGTGCATGATGATGATGATGATGATG 900
301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyValAlaThrThrLeuGluGly 320
901 AAGCCCATGTTAGTAGAGGGCAGCTTCGATGCAATTTGGTGTACCAACGCTTGAAGATAC 960
321 ArgLysThrValGluLysAspAlaAlaPheGluArgArgPheGlnValThrValAla 340
961 AGGAAATATGTTGAGAAAGATGCTGCTTTGAGAGGAGGTTCCAAACAGTCTATGTCGG 1020
341 GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysThrGluGly 360

1021 GAGCCAGTGTGCTGTGACACCATTTAGTATCTTAGAGGACTCAAGGAGAGATATGAGGGA 1080
361 HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg 380
1081 CATCATGTTGTGCGAATCCACAGACAGAGCTCTTATAAATGCTGCTCAGCTGTCTGCTCGT 1140
381 TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla 400
1141 TACATACTGCTGGCATTACCGGATTAAGCAATTCATTTGTTGATGAGGCTTTGTGGC 1200
401 AsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArg 420
1201 AATGTGAGAGTCCAGCTTGATAGTCAACCTGAAGAGATTGATAACCTTGAAGAGGAGG 1260
421 MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla 440
1261 ATGCACTGGGAAATGGAACCTTCAGCCTTGGAAAGGAGAGGATAAAGCCAGCAAGCT 1320
441 ArgLeuIleGluValArgLysGluLeuAspLeuArgAspLysLeuGlnProLeuThr 460
1321 CGACTTATAGAGTTCGGAAAGAGCTTGTGACTTGAGACCTGAGACAGCTTTCAGCTCTCAG 1380
461 MetLysThrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArg 480
1381 ATGAATACAGAAAGGAGAAAGAGAGATTCAGAGATTTCGAAGGCTTAAACAGAAAGA 1440
481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgGlyrAspLeuAlaArgAlaAla 500
1441 GAAGAGCTCATGTTTTCTTCGAGGAGGAGAGAGAGATGATGACCTTGCACAGAGCTGCT 1500
501 AspLeuArgThrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr 520
1501 GATCTAAGATATGCGCAATTCAGAGTGAATCTGCAATTCGCCAATTCGAAGGAACT 1560
521 SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluVal 540
1561 TCTTCTGAAGAGATGTGATGCTCACAGAAAGCTTGGGGCTTGAACACATTCGTGAGGTT 1620
541 ValSerArgThrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeu 560
1621 GTGAGCGCTTGACAGGGATTCAGTACGAGACTTGGCCAAATGAGAAAGGAGAGTTG 1680
561 IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAla 580
1681 ATTGCTCTGCTGATAGTTGTCATAGCGGTTGTGGACAGAGTCAAGCGTAAATGCA 1740
581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnProThrGly 600
1741 GTTCTGAGGCAATTCCTAAGGTCAGGCGAGGACTTGGAGGCGCCACACAGCCAACTGGA 1800
601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
1801 TCATCTTATTCCTTGACCACTGCTGTGGCAAACTGAGCTGCCAAGGCTCTTGTCT 1860
621 GluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluThrMetGlu 640
1861 GAGCAGCTGTTGATGATGAAACCTCTTAGTTCGATTCGATTCGATTCGATTCGATTCGAA 1920
641 GlnHisSerValSerArgLeuIleGlyAlaProProGlyThrValGlyHisGluGluGly 560
1921 CAACACTCTGCTCTGCTGCTCAATGGGGCACCCAGGGGTATGTTGTCACAGGAGG 1980
661 GlyGlnLeuThrGluAlaValArgArgProThrCysValIleLeuPheAspGluVal 680
1981 GGACCACTAATGAGGCTGTGAGGAGCGAGCTTATTTGTCATACCTCTTTGATGAAGTG 2040
681 GluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGlyArgLeu 700
2041 GAGAAAGCTCATGTTGCTGCTCAACACTCTGCTCAAGTTCGATTCGATTCGATTCGATTCG 2100
701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSerAsn 720

2101 ACAGCGGGCAGGACGAGCAGTCGATTTCAGGAACCTCGGTGATATATCATGACATCAAC 2160
 721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740
 2161 CTGTGCTGACACCTCTCTTGAGGGCTAACTGGGAATACATGAGAGTGGCCGG 2220
 741 AspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAenArgLeuAsp 760
 2221 GACTGTGTGATGCGGAGGTGAGGAACACTTCAGACCAAGAGCTCTTGAACAGGCTTGAC 2280
 761 GluIleValValPheAspProLeuSerHisAspClnLeuArgLysValAlaArgLeuGln 780
 2281 GAGATTGTGTGTCGACCCCTTCATGACACAGTGTGAGGAAGTAGCTCGGCTCAA 2340
 781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800
 2341 ATGAAGACGCTTGCTGTCGGCTTGCTGAAAGAGGAGTTCCTTTGGCAGTCACTGATGCT 2400
 801 AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820
 2401 GCTTTGACTATATCTTTGGCAGAGATTATGACCCGGTGTATGCTGCTAGGCCCTATAAGG 2460
 821 ArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIle 840
 2461 AGATGATGAGGAAGAGGTGTCGACAGAACTGTCAAGAGATGTTGTCGTGAGGAATC 2520
 841 AspGluAenSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu 860
 2521 GATGAAGAACTCCACTGTTATAGATGACAGGCGCTGCTGATCTTGTGTACCGGGTAGAA 2580
 861 SerGlyLysLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla 880
 2581 AGTGGAGGTCTAGTGGACGCTTCAACAGGCAAGATCAGATGCTGCTGATTATATGCT 2640
 881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIle 900
 2641 AACGGCCAAAGCAAGTATGACGCTCAGCGGTGAGAGATGAGATCGAGGAATA 2700
 901 GluAspAspAsnGluGluMetIleGluAsp 911
 2701 GAGATGACGATAATGAGGAATGATCGAGAT 2733

ILT 3

ABZ42033

27-FEB-2003 standard; cDNA; 2736 BP.

ABZ42033;

27-FEB-2003 (first entry)

Arabidopsis thaliana gene #17 modulated by PTGS.

Posttranscriptional gene silencing; PTGS; plant; transformation; gene; ss.

Arabidopsis thaliana.

Key Location/Qualifiers
 CDS 1..2736
 /*tag= a

W0200281695-A2.

17-OCT-2002.

05-APR-2002; 2002WO-EP03806.

06-APR-2001; 2001US-282049P.

(SYGN) SYNGENTA PARTICIPATIONS AG.
 (FRIE-) FRIEDRICH MIESCHER INST.

Zhu T, Glazov EA, Meins F, Wang X, Chang H;

XX WPI; 2003-103337/09.
 DR P-PSDB; ABP81189.
 XX
 PT Novel polynucleic acid segment useful for modulating gene expression
 within a cell by posttranscriptional gene silencing, and for augmenting
 a plant cell genome -
 XX
 PS Claim 18; Page 185-186; 438pp; English.
 XX
 CC The invention relates to a novel isolated polynucleic acid segment
 modulated within a cell by posttranscriptional gene silencing (PTGS). The
 CC invention specifically relates to a method to identify an expression
 CC product that is modulated by PTGS. The polynucleotide is useful for
 CC modulating the gene expression within a cell by PTGS, by introducing the
 CC polynucleic acid into a cell and expressing the nucleic acid segment in
 CC the cell to form a product. The polynucleic acid segment is also useful
 CC for augmenting a cell genome, and for augmenting a plant genome, by
 CC contacting a plant cell with the segment to produce a transformed plant
 CC cell, and growing the transformed plant cell to produce a differentiated
 CC transformed plant. The sequences shown in ABZ42017 - ABZ42142 represent
 CC segments of A. thaliana cDNA modulated by PTGS.
 XX
 SQ Sequence 2736 BP; 784 A; 481 C; 785 G; 686 T; 0 other;

Alignment Scores:
 Pred. No.: 0 Length: 2736
 Score: 4545.00 Matches: 910
 Percent Similarity: 99.89% Conservative: 0
 Best Local Similarity: 99.89% Mismatches: 1
 Query Match: 99.89% Indels: 0
 DB: 25 Gaps: 0

US-09-812-350-17 (1-911) x ABZ42033 (1-2736)

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 Db 1 ATGAATCCAGAGAAATTCACACAAAGCAACACGAGACAATTGCTACAGCTCATGAGCTA 60
 QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 Db 61 GCTGTGATGTCAGGACATGCTCAATTCCTCTTGTGATTTAGTGTGCTTGTATCTCT 120
 QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
 Db 121 GATCCACCGGTATATTTCTCAAGCAATCTCTAGTCCGGTGGCGAGAACGAGCTCAA 180
 QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
 Db 181 TCTGCTGAAGAGTGTATCAATCAAGCCTTGAAGAGCTTCTTCAATCTCTCCACCT 240
 QY 81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgAlaGlnAlaAlaGln 100
 Db 241 GATGATATTCAGCGAGTCTTAGTCTTATTAAGTCAATTCCTGCTCAAGCTGCTCAG 300
 QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 120
 Db 301 AAGTCACGAGGTGATCACTCAATTTGGCTGTTCACCACTGATATATGGGTCTTCTTGAAGAT 360
 QY 121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
 Db 361 TCTCAATCAGCGATTTTGTGAACGAAGTCCGCTGTAGCGAGCGGAGGTTAAGTCTGAG 420
 QY 141 ValGluLysLeuArgGlyLysGluLysLysValGluSerAlaSerGlyAspThrAsn 160
 Db 421 GTTGAGAGCTTCTGGGAAGAGAGGAGAAAGTTGAGAGTCTTTCAGGGGACACAAAT 480
 QY 161 PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
 Db 481 TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGAGGAGCTTGATCT 540
 QY 181 VallIleGlyArgAspGluLeuIleArgArgValValArgIleLeuSerArgArgThrLys 200

541 GTGATTGGTCGTGATGAGGAGATTAGAAAGAGTCGTGAGGATTCTTTCCGAGGAGAACGAAG 600
201 AsnAspProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220
||||| RACAACTCTGTCTTATTGGAGAGCCAGAGTGTGTAAACAGCTGTGTGTAAGGTTTA 660
221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
661 GCACAAAGGATTGTAAAGAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAAATTCG 720
241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGluArgLeu 260
721 TTGGACATGGGTGGTGTAGTCTGTGTCTTAATACCGAGGAGAGATTGCAAGAAAGGTTG 780
261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIlePheIleAspGlu 280
781 AAATCTGTTTGAAGAGAGTTCAGGAGCGCTGAAGCGCAAGTGAATCTCTTATTGATGAG 840
281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300
841 ATTCATTGTGTTCTGGTCTGGCAAACTGAAGGTCGATGGATGCAGCTAAATCTGTTTC 900
301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGluTyr 320
901 AAGCCCACTGTAGTGTAGAGGCGAGCTTCGATGCATTGGTGCTTACAACTTGAAGAAATAC 960
321 ArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyrValAla 340
961 AGCAAAATATGTTGAGAAAGATGCTGCTTTGAGAGAGAGGTTCCAAACAGTCTATGTTGG 1020
341 GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGly 360
1021 GAGCCAAAGTGTGCTGACACCAATTAGTATCTTTAGAGGACTCAAGGAGAGATGATGAGGA 1080
361 HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg 380
1081 CATCATGTTGTGGATCCAGACAGAGCTCTTAAATGCTGTCTGCTGCTGCTGCTGCTGCT 1140
381 TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla 400
1141 TACATAACTGGTTCGGCATTTACCGGATAAAGCAATTTGATTTGTTGATGAGGCTTGTGCG 1200
401 AsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArg 420
1201 AATGTGAGAGTCCAGCTTCATAGTCAACCTGGAAGAGATTGATAACCTTGAAGAGAGAGG 1260
421 MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGlyLysAspLysAlaSerLysAla 440
1261 ATGCAGCTGGAATTTGAATTTCAAGCTTTGAAAGGAGAGATTAAGCCAGCAAGCT 1320
441 ArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThr 460
1321 CGACTTATAGAGTGGGAAAGAGCTTGTATGACCTGATGACCTGAGAGACAAGCTTCTCAG 1380
461 MetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArg 480
1381 ATGAAATAAGAGAGAGAGAGAGAGATTTGATGAGATTGGAAGGCTTAAACAGAAAAAGA 1440
481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAla 500
1441 GAAGAGCTCATGTTTCTTGTGAGGAGGAGAGAGATGACCTTGCAGAGAGCTGCT 1500
501 AspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr 520
1501 GATCTAAGATATGGCGCAATTTCAAGAGTGGATCTGCAATTTGCCCAACTTGAAGAACT 1560
521 SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluVal 540
1561 TCTTCTGAAGAGAAATGTGATGCTTCAAGAAAAAGTTGGGCCCTGAAACACATCTGAGGTT 1620
541 ValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeu 560
1621 GTGAGCGGTTGGACAGGATTCAGTGAAGAGACTTGGCCAAATGATGAGAGAGAGGTTG 1680

561 IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAla 580
1681 ATTGGTCTCTGATAGTTGTCATAGCGGTTTGGGACAGATCAAGCGGTAAATGCA 1740
581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
1741 GTTCTCAGGCAATTTCAAGTCAAGGGCAGGACTTTGGAAGGCCACACAGCAACTGGA 1800
601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
1801 TCATCTTATTTCTTGGACCACTGTTGTTGGCAAACTGAGCTCGCCAGAGCTCTTGCT 1860
621 GluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluLysMetGlu 640
1861 GAGCAGCTGTTTGTATGATGAAAACCTCTTAGTTCGATTCGATATCGGAATATATGGA 1920
641 GlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGluGluGly 660
1921 CAACACTCTGCTCTCGCTTCATTTGGGGCACCAAGGCTATGTTGTCACGAGGAAGT 1980
661 GlyGlnLeuThrGluAlaValArgArgArgProTyrCysValIleLeuPheAspGluVal 680
1981 GGAACAATACTGAGGCTGTGAGAGGGGACCTTATTTGTCATCTCTTTGATGAAGTG 2040
681 GluLysAlaHisValAlaValPheAsnThrLeuGlnValLeuAspAspGlyArgLeu 700
2041 GAGAAGGCTCATGTTGCTGCTTCAACACTCTGCTCAAGTTTGGATGATGTCGATTG 2100
701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSerAsn 720
2101 ACAGACGGGCAAGGAGGAGCAGTCGATTTTCAGAACTCGGTGATATCATGACATCAAC 2160
721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740
2161 CTTGCTGCTGAAACCTCTCTTGCAGGGCTAACTGGGAAAGTAACAATGGAAGTGGCCG 2220
741 AspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAsp 760
2221 GACTGTGTGATGCGGAGGTGAGGAAACACTTCAGACCAGAGCTCTTGAACAGGCTTGAC 2280
761 GluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGln 780
2281 GAGATTGTGTGTTCAGCCCTTTTACATGACACAGTCAGTTGAGGAAAGTAGCTCGGCTCA 2340
781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800
2341 ATGAAAGAGCTTGTCTGCTCCGCTTGTCTGAAAGAGAGTGTCTTTGGCAGTCACTGATG 2400
801 AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820
2401 GCTTTGACTATATCTTTGGCAGAGAGTTATGACCCGCTGTATGCTGTAGGCTTAAAG 2460
821 ArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIle 840
2461 AGATGGATGAGAGAGAGGTGGTGCAGAACTGTCTCAAGATGTTGTGTGAGGAAATC 2520
841 AspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu 860
2521 GATGAAACCTCCACTGTTTACATAGATGACGCGCTGGTGTATCTTGTGTACCGGTAGAA 2580
861 SerGlyGlyLeuValAspAlaSerThrGlyLysSerAspValLeuIleHisIleAla 880
2581 AGTGGAGGCTTAGTGGACGCTTCAACAGGCAAGAGTCAGATGTGCTGATTCATATTGCT 2640
881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysMetArgIleGluGluIle 900
2641 AACGGGCAAGAGAGAGTGTGATGCTCAGCGCTCAGCGGTGAGAGATGAGGATCGAGGAATA 2700
901 GluAspAspAsnGluGluMetIleGluAsp 911
||||| GAAGATCAGCAATAATGAGGAAATGATCGAGGAT 2733

ULT 4

66066

AAI66066 standard; DNA; 6376 BP.

AAI66066;

14-JAN-2002 (first entry)

Arabidopsis thaliana heat shock protein 101 gene (GenBank: AF218796).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton; canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato; Arabidopsis thaliana; ds.

Arabidopsis thaliana.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

P-PSDB; AAM51659.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAM51651-AAM51671). The transgenic plant has increased stress tolerance, especially to heat. The transgenic plant has grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 6376 BP; 1914 A; 1150 C; 1470 G; 1842 T; 0 other;

gment Scores:

d. No.:	1.31e-315	Length:	6376
Matches:	4388.00	Matches:	908
cent Similarity:	86.56%	Conservative:	0
t Local Similarity:	86.56%	Mismatches:	3
ry Match:	96.44%	Indels:	140
	22	Gaps:	4

09-812-350-17 (1-911) x AAI66066 (1-6376)

1 MetAnProGluLysPheThrHisLysThrAnGluThrIleAlaThrAlaHisGluLeu 20

1132 ATGAATCAGAGAAATTCACACACAGACAAACAGACAAATGTCTACAGCTCATGAGCTA 1191

21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLeuSer 40

1192 GCTGTGAATGAGGACATGCTCAATTCCTCTTGCAATTTAGCTGGTCTTTCATCTCT 1251

Qy	41	AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGluAsnAlaGln 60
Db	1252	GATCCACCGGTATTTCTCAAGCAATCTTAGTCCCGTGGAGAACGCGCTCAA 1311
Qy	61	SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
Db	1312	TCTGCTGAAGAGTGATCAATCAAGCTTGAAGAGCTTCTTCACAATCTCTCCACCT 1371
Qy	81	AspAspIleProIleSerSerSerLeuLysValIleArgAlaGlnAlaGln 100
Db	1372	GATGATATCCACGAGTCTAGTCTTATTAAAGTCAATTCGTCTCAAGCTGTCTAG 1431
Qy	101	LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 120
Db	1432	AGTCAAGAGTGATCTACTCTTTGGCTTTGACAGCTTGATTATGGGTCTTCTTGAGAT 1491
Qy	121	SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
Db	1492	TCTCAATCAGGAGTTGTTGAACGAAGTCGGTGTAGCGACGCGAGGGTAAAGTCTGAG 1551
Qy	141	ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
Db	1552	GTTGAGAAGCTTCGTGGAAAGAAAGGAAAGATTGAGAGTCTTTCAGGGGACACAAAT 1611
Qy	161	PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluAlaGlyLysLeuAspPro 180
Db	1612	TTTCAAGCTTTAAAGACTTATGGAAGAGATTGGTTGACGACGAGGAGCTTGATCTCT 1671
Qy	181	ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgThrLys 200
Db	1672	GTGATTGCTGCTGATGAGGAGATTAGAGAGTCTGAGAGTCTTTCGAGGAGAACGAAG 1731
Qy	201	AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220
Db	1732	AACATCTCTGCTGCTTATGAGAGAGCAGAGTTGGTAAACAGCTGTGTTGAAGGTTTA 1791
Qy	221	AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
Db	1792	GCACAAAGGATTGTGAAGAGAGATGTGCCACACAGCTTACTGATCTGAGATTAATTCG 1851
Qy	241	LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGluArgLeu 260
Db	1852	TTGGACATGGTGGTGTAGTGTGCTGTGTAAATACCGAGGAGAGTTTCAAGAAAGGTTG 1911
Qy	261	LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
Db	1912	AAATCTGTTTGAAGAGAGTTGAGGACGCTGAAGGCAAGTGAATCTCTTTATTGATGAG 1971
Qy	281	IleHisLeuValLeuGlyAlaGlyLysThrGlySerMetAspAlaAlaAsnLeuPhe 300
Db	1972	ATTCAATTTGTTCTTGTGCTGCGCAAACTGAAGGGTCCATGGATGCAGCTAATCTGTTTC 2031
Qy	301	LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGluTyr 320
Db	2032	AAGCCCATGTTAGCTAGAGGCGAGCTTCGATTCATTTGTTGTCTACACCGCTTGAAGATAC 2091
Qy	321	ArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnValTyrValAla 340
Db	2092	AGGAATATGTTGAGAAAGATGCTGCTTTGAGAGAGGTTTCAACAGTCTATGTTGG 2151
Qy	341	GluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlyLysTyrGluGly 360
Db	2152	GAGCCAAGTGTGCTGACACCATTAGTATCTCTTAGAGGACTCAAGGAGAGATATGAGGGA 2211
Qy	361	HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg 380
Db	2212	CATCATGTTGCGAATCCACAGACAGCTCTTATAATGCTGCTCAGCTGCTGCTGCT 2271
Qy	381	TyrIleThr----- 383
Db	2272	TACATAACTGGTATGTTAAGATCTTAACTTAAAGCTGATGTTTATGGTTTTCATATAGTG 2331
Qy	384	-----GlyArgHisLeuProAspLysAlaIle 392

2332 TTGTTCTTGACATGACATTTTGTGGCGATTATAGTTCGGCATTTTACCGGATAAGCAAT 2391
 393 AspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGlu 412
 2392 GATTGGTTGATGAGGCTTGTGGAATGTGAGAGTCAGCTTATAGTCAACCTGAAGAG 2451
 413 IleAspAsnLeuGluArgLysArgMetGlnLeuGluLeuGluLeuHisAlaLeuGluArg 432
 2452 ATTGATAAACCCTTGAAGGAAGAGATGATGAGCTGGAAATTTGAACTTCACGCTTGGAAAG 2511
 433 GluLysAspLysAlaSerLysAlaArgLeuLeuGluVal 445
 2512 GAGAAGGATAAAGCCAGCAAGCTCGACCTTATAGAGGT-ATGATACAGCTTTCTTTCTC 2570
 445 ----- 445
 2571 ATCAGTAGATAGCGCTTCAATATAAGCATATTGTGCTTGTCTGGAAATCTTGATGTT 2630
 445 ----- 445
 2631 AGCATCTGTTTTGCGATGAGCTTTTGATATAATTTCTCAAACTATGAATCTGTAGTG 2690
 446 ArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThrMetLysTyrArgLys 465
 2691 CGGAAGAGCTTGATGACCTGAGAGACAAGCTTCAGCTCTCAGATGAATAACAGAAAG 2750
 466 GluLysGluArgLysAspGluLeuArgLysGlnLysArgGluGluLeuMetPhe 485
 2751 GAGAAACAGAGATTGATGATTCGAGATTCGAGGCTTAACAGAAAGAGAGAGCTCATGTT 2810
 486 SerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAlaAspLeuArgTyrGly 505
 2811 TCTTTGAGGAGGAGAGAGAGATATGACCTTCAGAGCTGCTGATCAAGATATGCG 2870
 506 AlaTleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAsn 525
 2871 GCAATTCAGAAAGTGAATCTGCAATTTGCCCACTTGAAGAACTTCTTCTGAAGAGAAT 2930
 526 ValMetLeuThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThr 545
 2931 GTGATGCTCACAGAAACGTTGGCCCTGAACATATGCTGAGCTTGTGAGCCCTTGACA 2990
 546 GlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeuLeuAlaAsp 565
 2991 GGGATTCAGTGCAGACTTGGCCAAATGAGAAGGAGAGGTTGATGCTCTGCTGAT 3050
 566 ArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIle 585
 3051 AGTTGCTAAGCGGTTTGGGACAGATCAAGCGTAAATGCAATTCAGTTCTGAGGCAAT 3110
 586 LeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeu 605
 3111 CTAAGGTCAAGGCGAGACTTGAAGGCCACACAGCAACTGATCATTTCTTTATTCCTT 3170
 606 GlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAsp 625
 3171 GSAACCAACTGCTGTGTGCAAACTGAGCTCGCAAGGCTCTTCTGAGCAGCTGTTTGTAT 3230
 626 AspGluAsnLeuLeuValArgLysAspMetSerGluTyrMetGluGlnHisSerValSer 645
 3231 GATGAAACCTCTTAGTTCCGATTTGATATGCGAATATATGGAACAACACTCTGTCTCT 3290
 646 ArgLeuIleGlyAlaProProGly 653
 3291 CGGCTCATTTGGGCAACCACCAGGAGTAAGTAATTTATCTTTTCAATATCTTCTCA 3350
 654 ----- 654
 3351 TTATGAAATATTTTCTTACTTACACGATTACTACCTTTCTCAGGTATGTTGTCAC 3410
 658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe 677

Db 3411 GAGGAAGGTGACAACTAACTAGGCTGTGAGGAGCGACCTTATTGTGTCACTCTTT 3470
 QY 678 AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp 697
 Db 3471 GATGAAGTGGGAAGGCTCATGTTGCTGTCTTCAACACTCTGCTCCAAGTTTGGATGAT 3530
 QY 698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMet 717
 Db 3531 GGTGATTTGACAGACGGCAAGGAGGACAGTCGATTTTCAGAACTCGGTGATATCATG 3590
 QY 718 ThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGlu 737
 Db 3591 ACATCAAACTTGGTGTGTAACACCTCTTCAGGGCTAACTGGGAAAAGTAACAATGAA 3650
 QY 738 ValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsn 757
 Db 3651 GTGGCCCGGACTGTGTGATGCGGAGGTGAGGAAACACTTCAGACCAGAGCTCTTGAAC 3710
 QY 758 ArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAla 777
 Db 3711 AGGCTTGACGAGATTGTGCTTGCACCCCTTTCAATGACAGCTTGAGGAAAGTAGCT 3770
 QY 778 ArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaVal 797
 Db 3771 CGGCTTCAAAATGAAAGACGTTGCTGCGGCTTCTGAAAGAGAGTTGCTTTGGCAGTC 3830
 QY 798 ThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProVal 813
 Db 3831 ACTGATGCTGCTTTGGACTATATCTTGGCAGAGAGTTATGACCCGGT-AACTCTCTTCAC 3889
 QY 813 ----- 813
 Db 3890 ATGATGACAGAATGCATATCATTTGTGATCTCTCAGGCTGGTTTGTCTGTGACTGACTC 3949
 QY 814 -----TyrGlyAlaArgProIleAraArgTyr 822
 Db 3950 TTGTGGTTGCTTTCTGTTTGTGGTTAGTGTGATGCTGCTAGGCTTATAGGAGATGG 4009
 QY 823 MetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIleAspGlu 842
 Db 4010 ATGGAGAAGAAGTGTGACAGAACTCTCAAGATGTTGTGCTGAGGAAATCGATGAA 4069
 QY 843 AsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGluSerGly 862
 Db 4070 AACTCCACTGTTTACATAGATGACGGGCTGCTGATCTTGTGTACCGGTAGAAAGTGA 4129
 QY 863 GlyLeuValAspAlaSerThrGlyLysSerAspValLeuIleHisIleAlaAsnGly 882
 Db 4130 GGTCTAGTGGACGCTTCAACAGGCAAGATCAGATGTGCTGATTCATATTGCTTAACGGG 4189
 QY 883 ProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIleGluAsp 902
 Db 4190 CCAAAGAGAGTGTGATGCTCAGCTCAGCGGTGAAGAAGATGAGGATCGAGGAAATAGAAGAT 4249
 QY 903 AspAspAsnGluGluMetIleGluAsp 911
 Db 4250 GACCATATGAGAAATGATCGAGGAT 4276
 RESULT 5
 ID AAI66067 standard; DNA; 3049 BP.
 AC AAI66067;
 XX
 DT 14-JAN-2002 (first entry)
 XX
 DE Glycine max heat shock protein gene (GenBank: L35272).
 KW Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 KW Arabidopsis thaliana; ds.
 OS Glycine max.

```

1 MetAsnProGlnLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
136 ATGAATCCTCAGAGAAGTTTACTCAACAAGCTAATGAAGCTCTTGCTCGGCCACGAGCTC 195
21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
196 GCGATGAGTTTCAGCCACGGCAATTTGATCTCCATCTCACTTGCCGCCACGGCTGATTTC 255
41 AspProThrGlyIlePheProGlnAlaIleSerSerAla--GlyGlyGluAsnAlaAla 59
256 GATCCCAACGGCATCTTCGTGAGCGATATACACGGCGGGCGGGAGAAATCGGCA 315
60 GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 79
316 CGCGCGGTGGAGCGAGTGTTGAACACGAGCTCTGACAGAGCTACCTCGCCAGTCCCTCCG 375
80 ProAspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAla 99
376 CCGGACGAGGTCCGGCGGACCAACCTCGTGAGGGCCATCAGGAGAGCACACGGCGGCG 435
100 GlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuGlu 119
436 CAAAAATCACTGGCGACACGCGTTTGGCGGTGATCACTGATCTCTCGAATCCTCGAA 495

```

480 ArgGluLeuMetPheSerLeuGlnAlaGluArgGlyTyrAspLeuAlaArgAla 499
1576 AGAAGAGAGCTTCTTTTGTTCACAGAGCTGAGAGAAGATATGATCTGGCTAGAGCT 1635
500 AlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGly 519
1636 GCACACCTGGCATATGAGCAATTCAGAGGTTGGAACCTCAATACAACTTGAAGGG 1695
520 ThrSerSerGluGluAanValMetLeuThrGluAanValGlyProGluHisIleAlaGlu 539
1696 ---AGCACTGAAGAGAATCTGATGTTGACCTGAACCTGTTGACCGGAGCAATATGCTGAG 1752
540 ValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAenGluLysGluArg 559
1753 GTTGTGAGCGCTGACCGGTATACCACTTACAGGCTTGCCCAATGAAAGAAAGA 1812
560 LeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAenGlnAlaValAan 579
1813 TTGATTTGGACTTGTGTGACAGATTGCACAGCAGAGTTGTAGCAAGACCAAGCTGTTAAT 1872
580 AlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThr 599
1873 GCTGTCTGAGGCTGTGTTGAGATCTAGAGCTGGGCTAGGAAGACCTCAACAACT 1932
600 GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeu 619
1933 GGTTCATTCCTATTCCTTGTCCAACTGGTGTGGTAACTGAGCTTGCAGAGGCTCTT 1992
620 AlaGluGlnLeuPheAspGluAanLeuLeuValArgIleAspMetSerGluTyrMet 639
1993 GCTGAGCAACTATTGTGATAATGAAACCACTGCTGTGAGAATTGATATGTCGAATACATG 2052
640 GluGlnHisSerValSerArgLeuIleGlyValaProGlyTyrValGlyHisGluGlu 659
2053 GAGCAGCACTCGTTCCTCAAGATTGATGTGTGCACCAAGGATGTTGTGACATGAGAA 2112
660 GlyGlyGlnLeuThrGluAlaValArgArgProGlyCysValIleLeuPheAspGlu 679
2113 GGTGCAACAACTAACTGAAGCGGTGAAGCGAAGACCTTACAGTGTGTGTCTTTTGATGA 2172
680 ValGluLysAlaHisValAlaValPheAenThrLeuGlnValLeuAspAspGlyArg 699
2173 GTGAAAGCAATACATCTGTGTTCAATACCTCTCTTCAAGTTTGGACGATGGAGG 2232
700 LeuThrAspGlyGlnGlyArgThrValAspPheArgAanSerValIleIleMetThrSer 719
2233 TTAAGTATGATGACAGAGCGGTACTGTGGACTTGAACACTGATTAATATCATGACCTCA 2292
720 AsnLeuGlyAlaGluHisLeuAlaGlyLeuThrGlyLysValThrMetGluValAla 739
2293 AATCTTGGAGCAGACACCTCTCTCAGTGGACTTTCAGGAAATGTACCATGCAAGTTGCT 2352
740 ArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAanArgLeu 759
2353 CGGATCGGTATGAGAGGAGGAGGAGGAGCAATTTAGGCGCAATTTGCTGAATGACATT 2412
760 AspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeu 779
2413 GATGAATTTGTGATTTGACCCCTTTCATGACCAATTTGAGGAGGTTGCAAGGCTA 2472
780 GlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAsp 799
2473 CAATGAAGGATCTAGTAGTCTCTGTCTGAGAAAGGAATTCCTTGGCAGTCACTGAT 2532
800 AlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIle 819
2533 GCAGCACTAGCATATATACCTTCCGAGAGCTATGATCCGCTGTATGCTGTGACCAATA 2592
820 ArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGlu 839
2593 AGGAGTGGCTTGAGAGAGAGGTTGTCACAGAGTTCTTAGAATGCTTGTAGAGAGGAG 2652
840 IleAspGluAanSerThrValTyrIleAspAlaGly-----AlaGlyAspLeuValTyr 857

2653 ATTGATGAGAAATTCACCGTTTACATAGATCTGACCAATGGGGCGAGTTGGTCTAC 2712
858 ArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeu 876
2713 CGTGAGAAAAGAAATGGAGGGTGTAACTCTACAACTGGGCGAGAGTCCGATATCTTG 2772
877 IleHisIleAlaAenGly-----ProLysArgSerAspAlaAlaGlnAlaValLysLys 894
2773 ATTCAATACCTTAATGGACATGCACCTAAA---ACTGATGCTGTTCAAGCAGTCAAGAAG 2829
895 MetArgIleGluGluIleGluAspAspAanGluGlu 907
2830 ATGAGATTGAGAAATTCATGATCATGAATGCAAGAG 2868

RESULT 6
AAI66068
ID AAI66068 standard; DNA; 3052 BP.
XX AAI66068;
XX AC AAI66068;
XX 14-JAN-2002 (first entry)
XX Nicotiana tobacum 101 kDa heat shock protein gene (GenBank: AF083343).
XX Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
XX canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
XX Arabidopsis thaliana; ds.
XX Nicotiana tobacum.
XX WO200170929-A2.
XX 27-SEP-2001.
XX 20-MAR-2001; 2001WO-US08836.
XX 20-MAR-2000; 2000US-190769P.
XX 18-APR-2000; 2000US-198116P.
XX (ARCH-) ARCH DEV CORP.
XX Lindquist S, Queitsch C, Vierling E;
XX WPI; 2001-639123/73.
XX P-PSDB; AAM51665.
XX Transgenic plants with improved heat stress tolerance, useful for
XX producing animal feed, oil and synthetic products -
XX Claim 4; Page -; 91pp; English.
XX The invention relates to a transgenic plant, comprising a genetic
XX construct comprising a promoter operatively linked to a nucleic acid
XX sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
XX family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
XX increased stress tolerance, especially to heat. The plant is a cereal,
XX grass, ornamental plant, crop plant, food plant, oil-producing plant, a
XX synthetic product-producing plant, an environmental waste absorbing
XX plant, an alcohol plant, a medicinal plant, a recreational plant and/or
XX an animal feed plant. In particular, the transgenic plant is cotton,
XX canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
XX Arabidopsis thaliana. The plants may be used to produce animal feed,
XX alcohol, crop, oil, medicine or a synthetic product.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained from GenBank using the Accession Number
XX reference provided in the specification.
XX Sequence 3052 BP; 898 A; 536 C; 836 G; 782 T; 0 other;
SQ

Alignment Scores: 1.69e-279 Length: 3052
Pred. No.: 3896.00 Matches: 774
Score:

cent Similarity: 93.19% Conservative: 75
 t Local Similarity: 84.96% Mismatches: 56
 ry Match: 22 Indels: 6
 Gaps: 5

09-812-350-17 (1-911) x AA166068 (1-3052)

1 MetAsnProGluValPheThrHisLysThrAnGluThrIleAlaThrAlaHisGluLeu 20
 101 ATGAATCTTGAAAAATTCACCACAGACTACAGAGGCCCTTGCTGGGGCACTCGAGCTA 160
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 161 GCATATCCGCGAGGCATCTCAATTTAGCCTCTGCATATGGCTGTGGCTTAATATCT 220
 41 AspProThrGlyPheProGlnAlaIleSerSerAlaGlyGly---GluAsnAlaAla 59
 221 GATCAATGTTGTTATTTTCGACAGCGATTTGTCATGCTGGTGGGAATGAAGAGTAGCT 280
 60 GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 79
 281 RAATTCAGTGGAGCGGTATTTGATCAAGCGATGAAGAGCTACCTTCTCAACACCGGCT 340
 80 ProAspAspIleProAlaSerSerLeuIleLysValIleArgArgAlaGlnAlaAla 99
 341 CCTGACGAATCCCACTAGCACTTCCCTTATCAAGGTGTTACGCCGAGCAATCGTGG 400
 100 GlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGlu 119
 401 CAGAACTCTCGTGGTGACAGCCATTTAGCAGTGGATCAGTTGATTTTAGGACTGCTAGAA 460
 120 AspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
 461 GACTCCCAATTCAGATCTTTGAAAGAGCTGGAGTGAGTGATCAAGAGTGAATCA 520
 140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159
 521 GAGGTAGAGAACTTAGAGGAAGGAAGGAAGAAAGTCCGAAGTGTCTCAGGGGACACC 580
 160 AsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAsp 179
 581 ACATTCCAAGCACTCAACACTTATGGCCGTGATCTTGTGGAACAGCAGGAAAGCTTGAT 640
 180 ProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArgThr 199
 641 CCCGTGATTTGGTGGATCAGAAATTAAGAGATCGTTCGGATTTTATCAAGAGGACT 700
 200 LysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGly 219
 701 AAGAACACCCGGTCTTATTTGGAGAGCCCGTGTGGTAAACAGCAGTGTGTTGAAGGG 760
 220 LeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIle 239
 761 CTAGACAGAGATGTGACGTGGTGTATTCACAGTAAATTTAGCTGATGTTAGGCTTATA 820
 240 SerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluArg 259
 821 GCATTTGATATGGAGCGCTAGTTGCTGGAGCTAAGTACAGAGTGAATTTGAAGAGAGG 880
 260 LeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAsp 279
 881 CTGAAGGCTGTGCTGAAGAAGATTTGAAGAGCGGAAGGAAAGTAAATCTTTCAITGAC 940
 280 GluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeu 299
 941 GAGATACATTTAGTCTCGTGGTGGTGGAGAGAGGCTATGGATGCTGCTTAATCTG 1000
 300 PheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGlu 319
 1001 TTTAAGCAATGCTAGCAGAGGTCATTTACGTTGCAATTTGTCACACTACACTCGAGGAG 1060
 320 TyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyrVal 339

Db 1061 TACAGGAAGTATGTTGAGAAGGANGCTGCATTTTGAGAGGCGTTCAGCAGGTGTATGTT 1120
 Qy 340 AlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGlu 359
 Db 1121 GCTGAGCCTAGTGTGCTGACACTATTAGTATTCTCCGTGGTGGTGAAGGAGAGTATGAA 1180
 Qy 360 GlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAla 379
 Db 1181 GGGCATCATGCTGTCAAAATTCAGACAGAGCTCTTGTAGTGGCTGCCAGCTCTCATCT 1240
 Qy 380 ArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCys 399
 Db 1241 CGGTACATTACAGCTGCACATCTGCCAGATAGAGCTATTGACCTAGTTGATGAAGCTTGT 1300
 Qy 400 AlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLys 419
 Db 1301 GCAAAATGTTAGATTGAGCTTCACTTACCTGAGGAAATTCACAATCTTGAGAGGAG 1360
 Qy 420 ArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLys 439
 Db 1361 AGAATTCAGCTAGAGTGAACCTTCACTCTCGAGAAGGAAACACAAAGCTAGCAAA 1420
 Qy 440 AlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeu 459
 Db 1421 GCAGCTCTCATAGAGTGAGGAAGAACTTGATGATTTGAGAGCAAACTTCAACCTTTG 1480
 Qy 460 ThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLys 479
 Db 1481 ATGATGAGGTACAAAGAAAGAAAGAGATAGATGAGCTCGCAGGCTCAAGCAAAAG 1540
 Qy 480 ArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAla 499
 Db 1541 CGCGATGAGCTCATCTATCTTATCAAGAAGCTGAAGAGGATATGATCTGGCGAGGCA 1600
 Qy 500 AlaAspLeuArgTyrGlyValIleGlnGluValGluSerAlaIleAlaGlnLeuGluGly 519
 Db 1601 GCAGATCTGAGATATGGGCAATTCAGAAAGTGGAAATTCGAATAGCAATCTTGAGAGT 1660
 Qy 520 ThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu 539
 Db 1661 ACCTCAGCT---GAAAGTACAATGCTTAAACAGAGACTGTGGTCTCTGATCAGATCGCGAA 1717
 Qy 540 ValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArg 559
 Db 1718 GTTGTGAGTGTGGACTGTATTCGGTCTCAAGGCTTGGTTCAGAAATGAGAAGAGAGAA 1777
 Qy 560 LeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsn 579
 Db 1778 CTGATTTGGTCTTGGCGATAGATTGCCAAAGAGTGGTTCGGGCAAGATCATCGAGTTAGA 1837
 Qy 580 AlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThr 599
 Db 1838 GCTGTTGCTGAAGCCGTGTTAAGGTCAGAGCTGGTGTAGGAAGGCCACAGCAACCACT 1897
 Qy 600 GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeu 619
 Db 1898 GGTTCATCTCTTTCTTGGGGCCAACTGGTGTGGAAGAGACAGAGCTCGCTAAAGCTCTT 1957
 Qy 620 AlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGluTyrMet 639
 Db 1958 GCAGAGCAGCTCTTTGATGATGATGATAAATCTGATGATCAGATAGACATGTCGAGTACATG 2017
 Qy 640 GluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHisGluGlu 659
 Db 2018 GAACAACACTCTGTTTCCGGCTGATGCTGCTCCACCGGTTATGTTGGGCATGATGAG 2077
 Qy 660 GlyGlyGlnLeuThrGluAlaValArgArgArgProTyrCysValIleLeuPheAspGlu 679
 Db 2078 GGAGGACAACTTACTGAAGCTGTAGGAGGGGGCTTACAGTGTGCTGCTATTTGATGAA 2137
 Qy 680 ValGluLysAlaHisAlaValPheAsnThrLeuLeuGlnValLeuAspAspGlyArg 699
 Db 2138 GTTGAGAAAGCCATCTCTGAGTGTGTTAATACATGCTTCAAGTCTTGATGATGAGG 2197

700 LeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleLeuMetThrSer 719
 2198 TTAACAGATGGTCAAGCGCCGACAGTTGATTTACCAACTCGTGATTTATGACTTCA 2257
 720 AsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLeuValMetGluValAla 739
 2258 AACTGGGAGCAGATATCTGTTGCTGCTGATTAATGGGCAATGTACCATGGAGACGCT 2317
 740 ArgAspCysValMetArgGluValAlaGlyHisPheArgProGluLeuLeuAsnArgLeu 759
 2318 CGTGAATCGTATCGACGAGGTGGAAACGAGTTTAAAGCCGCGACTCTGTAATCGGTG 2377
 760 AspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeu 779
 2378 GATGAGATTGTTGTTGATCTCTGTCACAGCAGTGTGAGCAAGTATGCGCTAC 2437
 780 GlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAsp 799
 2438 CAGATGAGGAGCGTTGCACCTACGCTGGCTGAGAGGGGTATTGCAATTGGCGCTTACTGAG 2497
 800 AlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyValaArgProIle 819
 2498 GCAGCTCCAGATGTCTACTACAGAGATTATGACCCGGTTTATGGTGCAAGACCTATT 2557
 820 ArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGlu 839
 2558 AGGAGATGTTGGAGAGGAAGTGGTGACCGAGTATCCAGATGCTGTGAAGGAGGAG 2617
 840 IleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGly-----AspLeuValTyr 857
 2618 ATTGATGAGAATCTCAACGGTTTACATAGATGCTGGGGTCAGCGGGAAGATCTTAACCTAC 2677
 858 ArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeu 876
 2678 AGGTGGAGAAAGTGGAGGGTGTGTAATGCTGCCACCGGCCAAATCTGATATG 2737
 877 IleHisIleAlaAsnGlyProLysArgSerAspAlaAlaGlnAlaValLysMetArg 896
 2738 ATTACGCTTCTTAATGGTCCC---AGGAGTGATGCTGTCCAGCAGTCAAGAAGATGAGG 2794
 897 IleGluGluIleGluAspAspAspGluGlu 907
 2795 ATTGAAGAAATTGAAGATGACGAAATGGAAGAT 2827

MULT 7

566070

AAI66070 standard; DNA; 3084 BP.

AAI66070;

14-JAN-2002 (first entry)

Zea mays heat shock protein HSP101 gene (GenBank: AF133840).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana; ds.

Zea mays.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

XX WPI; 2001-639123/73.
 DR P-PSDB; AAM51669.
 XX
 PT Transgenic plants with improved heat stress tolerance, useful for
 producing animal feed, oil and synthetic products -
 XX
 PS Claim 4; Page -: 91pp; English.
 XX
 CC The invention relates to a transgenic plant, comprising a genetic
 construct comprising a promoter operatively linked to a nucleic acid
 sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
 family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 increased stress tolerance, especially to heat. The plant is a cereal,
 grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 synthetic product-producing plant, an environmental waste absorbing
 plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 an animal feed plant. In particular, the transgenic plant is cotton,
 canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 Arabidopsis thaliana. The plants may be used to produce animal feed,
 alcohol, crop, oil, medicine or a synthetic product.
 CC Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained from GenBank using the Accession Number
 CC reference provided in the specification.
 XX

SQ Sequence 3084 BP; 670 A; 874 C; 1057 G; 483 T; 0 other;

Alignment Scores:

Pred. No.:	1.45e-278	Length:	3084
Score:	3883.50	Matches:	772
Percent Similarity:	92.88%	Conservative:	76
Best Local Similarity:	84.56%	Mismatches:	58
Query Match:	85.35%	Indels:	7
DB:	22	Gaps:	6

US-09-812-350-17 (1-911) x AAI66070 (1-3084)

Qy 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 Db 145 ATGAATCCGACAACTTCCACCAAGACGAGCGCATCGTGGGGCGCACGAGATT 204
 Qy 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 Db 205 GCGGTGGAGCGCGCCACGCGCAGCTCACGCCCTGCACCTGCCCGCATGTGGCTGCG 264
 Qy 41 AspProThrGlyIlePheProGlnAlaIleSerSerAla---GlyGlyGluAsnAlaAla 59
 Db 265 GACAGGGCGGCATCTCGCGCAGGCCATTCACGGGGCGCTCGGGGGCGCACGAGCGGCC 324
 Qy 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 Db 325 GGGGACTCGTTCAGAGCGGTGCTGAACAACTCGCTCAAGAAGCTGCGCAGTCCCG 384
 Qy 79 ProProAspAspIleProAlaSerSerSerLeuIleLysValIleArgAlaGlnAla 98
 Db 385 CCGCGGACTCTCGTCCCGGCTCCACGGCACTCATCAAGTCAATCCCGCGGCGAGTCC 444
 Qy 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 Db 445 GCGCAGAAAAACGGGGGACTCTCACCTGCCCTGACCACTGCTGCTGCTGCTGCTGCTC 504
 Qy 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
 Db 505 GAGGACTCGCAGATCTCCGACTGCCTCAAGGAGCGCGGTGTCCCGCGCGGGTGGCG 564
 Qy 139 SerGluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAsp 158
 Db 565 GCCGAGCTTCAGAAAGTCCCGCGGGGAGGGCGCGCTGGAGTCCCGCTCGGGGAG 624
 Qy 159 ThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeu 178
 Db 625 ACCAACTTCCAGGGCTCAAGACATACGCGCGGAGCTCTGTTGAGCAGCGCGGAGCTG 684

179 AspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArg 198
685 GACCCCGCTCATCGCCGCGACGAGAGATCCGCGCGTGGTGCATCTCTCGCGCGCG 744
199 ThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGlu 218
745 ACTAGAACACCCCGCTCTCATCGCGAGCCGCGGCTTGGCAAGACGCGCTCGTGGAG 804
219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
805 GGCCTCGCGACGCGATCTCGCGCGAGCTGCCAGTAACCTCTCGAGCTCGCGCTC 864
239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyArgGlyGluPheGlu 258
865 ATCGCGCTCGACATGCGCGCTCTCGTGGCGCGCGCAAGTACCGCGCGAGTTCGAGGAG 924
259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
925 CGGCTCAAGCGCTGCTCAAGAGGTGGAGAGCGCGAGCGGAGGTAATCTCTTCATC 984
279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsn 298
985 GACGAGATACCTCGTCTCGCGCGCGGCGAGCGAGGGTTCATGGACGCGCGCAAC 1044
299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
1045 CTGTTCAAGCCCAATGCTGGCGAGGCGACAGCTCAGGTGCATCGCGCGCACACGCTGGAG 1104
319 GluTyArgLysTyValGluLysAspAlaAlaPheGluArgArgPheGlnValTy 338
1105 GAGTACCGCAAGTAGTGAGAGAGACGCGAGGCTTCGAGCGCGGTTCCAGCAGGTGTT 1164
339 ValAlaGluProSerValProAspThrIleLeuArgGlyLeuLysGluLysTy 358
1165 GTCGCGAGCGCGAGCGTGCGCGACACCGTCACTTCTAGCGGACTCAAGGAGAGTAC 1224
359 GluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378
1225 GAGGGGCAACCATGGCGTGGAGTATCAGGACCGCGCGCTCGTGGCGCGACAGCTATCC 1284
379 AlaArgTyIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
1285 GCGAGGTACATCATGGTTCGGCACTCTGCTGCAAGCCATAGACTGTGGAGCGAGGCC 1344
399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuArg 418
1345 TCGCGCAATGTGAGGTGCGAGCTCGACAGCGCGAGGAGATTTGATACTCGAGAGG 1404
419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSer 438
1405 AAGAAATCCAGCTTGAAGTTGAGTCTCACCGCTCGAGAGGAGGAGGAGCAAGCGCCAGC 1464
439 LysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnPro 458
1465 AAAGCCCGCTGATCAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1524
459 LeuThrMetLysTyArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGln 478
1525 CTGACCATCAAGTACAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1584
479 LysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyArgPheLeuArg 498
1585 CCGCGCGAGGAGCTCAGTTCACTCCCTCGAGGAGGCGCGCGCGGAGGAGGAGGAGGAG 1644
499 AlaAlaAspLeuArgTyGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGlu 518
1645 GTGGCGGACCTCAAGTACGCGCGCTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1704
519 GlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAla 538
1705 ---AGCGAAACAGGGGAGAACCTGATGCTACCGGAAACCGTGGCGCGCTGGAACAAATTGCA 1761
539 GluValValSerArgTrpThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGlu 558

1762 GAGGTGCTGAGCGCTTGGACGGGTATTCCAGTGATCCCGGCTTGGCCAGAAACGACAGAG 1821
559 ArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaVal 578
1822 AGGTGCTGGCGCTTGGCTGACAGGCTTCCACAGAGGGTGGTGGCGCCACAGAGGGTGTG 1881
579 AsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnPro 598
1882 AGCCCGCTCGCAGAGGGCGGTGCTGAGGTCCAGGGCGCGTCTTGGCAGGCCACACAGCCC 1941
599 ThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAla 618
1942 ACTGGCTCGTCTCTGTTCTCTGGTCCGACTGGGTGGGAAAACCTGAGCTAGCAGCAGGCC 2001
619 LeuAlaGlnGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluTy 638
2002 CTAGCCGAAACAGCTGTTTCGACGACGAGAACCTTCTTGTCCGCATCGACATGTCGGAGTAC 2061
639 MetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyValGlyHisGlu 658
2062 ATGGAGCAGCACTCGGTGCTCGCTCATCGAGCACCGCTGGCTACGTCGGCCATGAA 2121
659 GluGlyGlyGlnLeuThrGluAlaValArgArgArgProTyCysValIleLeuPheAsp 678
2122 GAGGTGGCGAGCTGACTGAACAAGTGAGGAGGAGCGGTACAGCGGTGATCTCTGTTCCGAC 2181
679 GluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspGly 698
2182 GAGTTCAGAGGCGCCATGTCGCGCTTCAACACCTGCTCCAGGTCCTCGACGACGCGC 2241
699 ArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleMetThr 718
2242 AGTTGACGATGGCAAGCGACGAGCGGTGACTTCAGGAACACCGGTGATCATGACA 2301
719 SerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluVal 738
2302 TCGAACCTCGCGCGCGAGCACTCTCTGGTGGATGTTGGGGAAGAACTCCATGAAGGTC 2361
739 AlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArg 758
2362 GCTCGCATCTGTCATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2421
759 LeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArg 778
2422 CTCGACGAGATCTGATCTTCGATCTCTGTCACAGCAGCAGCTGAGGAGGAGTCTGCTCC 2481
779 LeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThr 798
2482 CTTGATGAAGGATGTCGCGCTTTCGCGAAAGGCGCATCTCTGCTGCTGCTGCTGCTG 2541
799 AspAlaAlaLeuAspTyIleLeuAlaGluSerTyArgProValTyArgLysAlaArgPro 818
2542 GACCGCATTTGACATCT 2601
819 IleArgTrpMetGluLysLysValValThrGluLeuSerLysMetValValArgGlu 838
2602 ATCAGAGGTGGATCGAGAGAGGGTGGTGGAGCAGCTCTCGAAGATGCTGATCCAGGAG 2661
839 GluIleAspGluAsnSerThrValTyIleAspAlaGlyAlaGly-----AspLeuVal 856
2662 GAGATCGACGAGAACTGCGCGTCTACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2721
857 TyArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspVal 875
2722 TACAGGTGGACCGGAGCGCGGTCTGGTGAACGCTGAGACCGGGGATGAAGTCGGACATC 2781
876 LeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMet 895
2782 CTGATCCAGTCCCAACAGCTCCACAGGAGGAGCGCTGCGCAGGCGCTCAAGAGATG 2841
896 ArgIle---GluIleGluAspAspAsnGluGlu 907

221 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
|||
:::

399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArg 418
 1294 TGCGCAATGTGAGGGTCACTTGACAGCCAGCTGAGAGATCGAACCTCGAGAG 1353
 419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSer 438
 1354 AAGAGATCCAGTGTGAAGTCCAGCTCCATGCCCTCGAGAGAGGAGGACCAAGGCTAGC 1413
 439 LysAlaArgLeuIleGluValArgLysGluLeuAspLeuAspLeuArgGlnPro 458
 1414 AAAGCTCGCTAGTGTGAGGAGAGAGTGTGACAGCTGAGAGACCAAGCTCGACCG 1473
 459 LeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgGluLysGln 478
 1474 CTGCAGATGAAGTATCGCAAGAGAGAGAGGATCGACGAGATCAGAGGCTGAACGAG 1533
 479 LysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArg 498
 1534 CGCGCGAAGAGCTGCAGTTCCTCTGAGGAGGCGGAGCGCGGATGATGCTGAGCTGGAG 1593
 499 AlaAlaAspLeuArgTyrGluValMetLeuThrGluAsnValGlyProGluHisIleAla 518
 1594 GTGCTGATCTCAGATCGGTGCTCTGAGAGGTTGACGCTGCCATTTGCTAAGCTGGAG 1653
 519 GlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAla 538
 1654 GGTGAGACCGGC--GAGAAITTCATGTAAACAGAGACTGTGCGCCCGGACCATGCT 1710
 539 GluValValSerArgTyrThrGlyLeuProValThrArgLeuGluGlnAsnGluLysGlu 558
 1711 GAGTGTGTGAGCCCTGGACTGCTGATTCCTGTCACCGAGGCTCGGACGAGATGAGAGGCG 1770
 559 ArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaVal 578
 1771 AGGCTGATCGGGCTGGCAGATCGATCATCAGAGGCTGCTGGACAGTATGAGGCGGCTC 1830
 579 AsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnPro 598
 1831 AATCAGTTGGAGAGGCTGTTCTACGCTCGAGGCTGCGCTTGGACCGCCGACGAGCT 1890
 599 ThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAla 618
 1891 ACTGTTTCATTCCTCTTGGACCGACCGGCTGTCGGAACCAACCGAGCTCGCAAGGCT 1950
 619 LeuAlaGluGlnLeuPheAspGluAsnLeuLeuValArgIleAspMetSerGluTyr 638
 1951 CTAGCTGAGCAGCTGTTGATGACGAGAACTGCTGCTCGCATCGACATGCTCGAATAC 2010
 639 MetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHisGlu 658
 2011 ATGAGCAGCATTCGGTTGCGCGCTAATCGAGCCCACTCGATATGTTGTCATGAA 2070
 659 GluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAsp 678
 2071 GAAGGGGGGAGCTGACCGAGCAAGTGAAGAGGAGGACATACAGTGTTCATCTTCGAC 2130
 679 GluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAspGly 698
 2131 GAGTTGAGAGGCGCATGTGGCGGTGTTCAACTCTGCTCAGGTCCTGGAGCATGGG 2190
 699 ArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleLeuMetThr 718
 2191 CGGTTGACCGAGCGGCAAGGAGGAGGCTGATTTAGGAACACGGTGTATCATCATGACC 2250
 719 SerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluVal 738
 2251 TCAAACCTTGGCGGAGACCTCTCTCGCGAATGCTGGGCAAT---TCGATGAAGGTT 2307
 739 AlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArg 758
 2308 GCTCGTATCTGTCATGACGAGGAGTGAAGGAGGATTTCCGCGGAGGCTGCTGAACCGT 2367

759 LeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArg 778
 2368 CTGACAGATGCTCATCTTCGACCTCTGTTCGATGACAGCTCGGAGGCTCGCTCGG 2427
 779 LeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThr 798
 2428 CTTACAGTGAAGATGTGCAGTCCGCTCTTCGAGAGGCGGCTTCTCTGCGCGTCACC 2487
 799 AspAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgPro 818
 2488 GAGCGCGCTCGACGTCATCTCTGCTCTTACGATCCGCTCTATGCGCCGACGCGCA 2547
 819 IleArgArgTyrMetGluLysValValThrGluLeuSerLysMetValValArgGlu 838
 2548 ATCCGAGATGATCGAGAGAGGATGATGACGAGCTCTCCAGATGTTGATCCGCGAG 2607
 839 GluIleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGly-----AspLeuVal 856
 2608 GAGATCGACGAGAACTCCACGCTGTACATCGACGCTGCGCCCGACCAAGTCCGACATC 2667
 857 TyrArgValGluSer---GlyGlyLeuValAlaSerThrGlyLysLysSerAspVal 875
 2668 TATGCGCTCGACCAAGCACGAGGCTGTGTAACGCGCGCCGACCAAGTCCGACATC 2727
 876 LeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMet 895
 2728 CTGATCCAGGTTCTTAGCGAGCTGTTGGGGCGGATGCGCGCACGCGCTGAGAGATG 2787
 896 ArgIle-----GluGluIleGluAspAspAsnGluGlu 907
 2788 AAGATCATCGAGCAGCAGCGAGGCTGGACGACGATCGAGGAA 2829

RESULT 9
 AA166075 standard; DNA; 2821 BP.
 AC AA166075;
 DT 14-JAN-2002 (first entry)
 XX
 DE Triticum aestivum 101kDa heat shock protein gene (GenBank: AF083344).
 KW Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 KW Arabidopsis thaliana; ds.
 XX
 OS Triticum aestivum.
 XX WO200170929-A2.
 XX 27-SEP-2001.
 XX 20-MAR-2001; 2001WO-US08836.
 XX 20-MAR-2000; 2000US-190769P.
 PR 18-APR-2000; 2000US-198116P.
 XX (ARCH-) ARCH DEV CORP.
 PA Lindquist S, Queitsch C, Vierling B;
 XX WPI; 2001-639123/73.
 DR P-PSDB; AAM51661.
 XX Transgenic plants with improved heat stress tolerance, useful for
 PT producing animal feed, oil and synthetic products -
 PS Claim 4; Page -; 91pp; English.
 XX The invention relates to a transgenic plant, comprising a genetic
 CC construct comprising a promoter operatively linked to a nucleic acid
 CC sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP)
 CC family amino acid sequence (AAM51651-AAM51671). The transgenic plant has

increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 2821 BP; 524 A; 902 C; 1035 G; 360 T; 0 other;

Alignment Scores:

Seq. No.:	1-12e-271	Length:	2821
Pre:	3750.00	Matches:	752
Percent Similarity:	91.09%	Conservative:	86
at Local Similarity:	81.74%	Mismatches:	70
ary Match:	83.30%	Indels:	12
	22	Gaps:	7

09-812-350-17 (1-911) x AAT66075 (1-2821)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 65 ATGAACCGGACAACTTCAGCGCAAGACCAACAGGCGCTGCTGGCGCGCACGAGCG 124
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLeuSer 40
 125 GCGTCGGAGCGCGCGCACGCGCAGATCAACCGCTTGACCTGCGCGCGCGCTGCGCGG 184
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaA--- 59
 185 GACAAGTCGGGCACTCTCGCGCAGGCGCTCGCGCGGCGTTCGCGGCGGAATGCTCGCGG 244
 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 245 GGGGACTCTGTCAGGCGCGTCTCGCGCGCGCGCTCAGGAAGCTGCGCTCGCAGTTCGCG 304
 79 ProProAspAspIleProAlaSerSerSerLeuLysValIleArgArgAlaGlnAla 98
 305 CCGCGGACTCTCGTGGCGCTCTCAAGGCGCTCATCAAGCCATTCGCGCGCGCAGTCTG 364
 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 365 GCGCAGAAGAAGCGCGGACTCGCACTCGCGCTCGACCGAGCTGCTCATGGGCTCTCTC 424
 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
 425 GAGGACCGCGAGATCGCCGCTCAAGAGGCGCGCGTGTCCGCGCTCGCGGGTGGCG 484
 139 SerGluValGluLysLeuArgGlyLysGlu---GlyLysLysValGluSerAlaSerGly 157
 485 GCGGAGCTCGACAGCTCCGCGCGGGACAACTCGCGCAAGTTCGAGTCCGCTTCGCG 544
 158 AspThrAsnPheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAlaGlyLys 177
 545 GACACCACTTCGAGGCGCTCAAGACGTACGCGCGGCACTCTGTCGAGGTGGCGCGCAAG 604
 178 LeuAspProValIleGlyArgAspGluGluLeuArgValValAlaArgIleLeuSerArg 197
 605 CTCGACCCCTCATCGGCGCGCAGAGAGATCCGCGCGCTGCTGCGCATCTCTCGCGC 664
 198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217
 665 CGCACCAAGAACACCCGCTCATCGCGGAGCGCGCTCGGCAAGCGCGCTCTGTC 724
 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 725 GAGGGCTTCGCGAGCGCGCTGTCGCGCGGAGCGTCCAGCAACCTGCTCGACGCTCCG 784
 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrglyGluPheGlu 257

Seq. No.:	1-12e-271	Length:	2821
Pre:	3750.00	Matches:	752
Percent Similarity:	91.09%	Conservative:	86
at Local Similarity:	81.74%	Mismatches:	70
ary Match:	83.30%	Indels:	12
	22	Gaps:	7

09-812-350-17 (1-911) x AAT66075 (1-2821)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 65 ATGAACCGGACAACTTCAGCGCAAGACCAACAGGCGCTGCTGGCGCGCACGAGCG 124
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLeuSer 40
 125 GCGTCGGAGCGCGCGCACGCGCAGATCAACCGCTTGACCTGCGCGCGCGCTGCGCGG 184
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaA--- 59
 185 GACAAGTCGGGCACTCTCGCGCAGGCGCTCGCGCGGCGTTCGCGGCGGAATGCTCGCGG 244
 60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 245 GGGGACTCTGTCAGGCGCGTCTCGCGCGCGCGCTCAGGAAGCTGCGCTCGCAGTTCGCG 304
 79 ProProAspAspIleProAlaSerSerSerLeuLysValIleArgArgAlaGlnAla 98
 305 CCGCGGACTCTCGTGGCGCTCTCAAGGCGCTCATCAAGCCATTCGCGCGCGCAGTCTG 364
 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 365 GCGCAGAAGAAGCGCGGACTCGCACTCGCGCTCGACCGAGCTGCTCATGGGCTCTCTC 424
 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
 425 GAGGACCGCGAGATCGCCGCTCAAGAGGCGCGCGTGTCCGCGCTCGCGGGTGGCG 484
 139 SerGluValGluLysLeuArgGlyLysGlu---GlyLysLysValGluSerAlaSerGly 157
 485 GCGGAGCTCGACAGCTCCGCGCGGGACAACTCGCGCAAGTTCGAGTCCGCTTCGCG 544
 158 AspThrAsnPheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAlaGlyLys 177
 545 GACACCACTTCGAGGCGCTCAAGACGTACGCGCGGCACTCTGTCGAGGTGGCGCGCAAG 604
 178 LeuAspProValIleGlyArgAspGluGluLeuArgValValAlaArgIleLeuSerArg 197
 605 CTCGACCCCTCATCGGCGCGCAGAGAGATCCGCGCGCTGCTGCGCATCTCTCGCGC 664
 198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217
 665 CGCACCAAGAACACCCGCTCATCGCGGAGCGCGCTCGGCAAGCGCGCTCTGTC 724
 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 725 GAGGGCTTCGCGAGCGCGCTGTCGCGCGGAGCGTCCAGCAACCTGCTCGACGCTCCG 784
 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrglyGluPheGlu 257

445	----	445
1815	TTGCTGTGAGTTCCTCGTTTGAACCTTTTGATACCTAACGTCGCTTAACTGTGGCCATTGTT	1874
446	-----ArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeuThrMetLysTyr	463
1875	CAGGTGAGGAAGAAATTGGACGATCTGAGGACAAGCTGCAGCCCTGACCATGAAGTAC	1934
464	ArgLysGluLysGluArgLysLeuAspGluLeuArgArgLeuLysGlnLysArgGluGluLeu	483
1935	AGGAAGGAGAAGAGAGAAATTGATGAGATCAGGAAGCTGAAGCAGCGCCGCGAGGAGCTC	1994
484	MetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAlaAspLeuArg	503
1995	CAGTTCACTCCCTGCAGAGGCCGAGCCGGATGAGACTGGCCCGTGTGGCCGACCTCAG	2054
504	TyrGlyAlaIleGlnValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGlu	523
2055	TACGGCGCCCTCCAGGAATTCGACGCTGCTATCTCCAAGCTGGAG--AGCGAAACACAGG	2111
524	GluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu	539
2112	GAGAACCTGATCTCACCGAAACCGTCGGCCCTGGAACAAATTGCAGAGGTATGTTATTAT	2171
539	-----	539
2172	TCCTGTTCACCGTCACAAAAATTTTCAGAGCAAGTCCCGAATTTCCGATCGCTCTCCTA	2231
540	-----ValValSerAr	543
2232	GTAGAGTAGTCGTGCGACGTCGTGAAATGGTGTTCGTCTATGGCAGGTGTGTGACCG	2291
543	gfrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLe	563
2292	TTGACGGGTATTCCAGTGCACCGGCTTGGCAGAACGACAGAGAGGCTGTTGGCCT	2351
563	uAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGl	583
2352	GGCTGACAGCTTCCACAGAGGTGGTCGCCACAGACAGAGGCTGTGAGCGCCCTCGACA	2411
583	uAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnProThrGlySerPheLe	603
2412	GGCGGTGCTGAGTGCAGGCGCGGTCTTGGCAGCCACACAGCCCATCGCTCGTCTCT	2471
603	uPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLe	623
2472	CTTCTGGGTCCGACTGGCTGGGGAAAATGAGCTGGCCACAGCGCTAGCCGCAACAGCT	2531
623	uPheAspAspGluAsnLeuValArgIleAspMetSerGluTyrMetGluGlnHisLe	643
2532	GTTTCGACGACGAGAACTTCTTGTCCGATTCGACATGTCCGAGTACATGAGCAGCCTC	2591
643	rValSerArgLeuIleGlyAlaProGly	653
2592	GGTTGCCCGCCTCATCGGAGCACCACCTGG-GTAACTAGCAGAAAAATGCATGCATCTTCG	2650
653	-----	653
2651	TATTTTAACTCGCAATTCGATGAACCTTGTTCGACAGTGGCGCGGACCTGTGACAG	2710
654	-TyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArgArgProTyrCy	673
2711	CTACGTCGCGCATGAAGAGGTGGGCAGCTGACTGNAACAAGTGAAGGAGGAGCCGTACAG	2770
673	sValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGl	693
2771	CGTGATCTCTTCGACGAGTTCGAGAAGGCCCATGTTCGCCGTTCACACCCCTGCTCCA	2830
693	nValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnse	713
2831	GGTCTTCGACGACGGCAGCTCAGCGATGGCAAGCAGGACGGTGGACTTCAGGAAC	2890
713	rValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLy	733

[illegible]

AAI66074:

14-JAN-2002 (first entry)

Triticum aestivum heat shock protein 101 gene (GenBank: AF097363).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton; canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato; *Arabidopsis thaliana*; ds.

Triticum aestivum.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.
18-APR-2000; 2000US-198116P.
(ARCH-) ARCH DEV CORP.
Lindquist S, Queitsch C, Vierling E;
WPI; 2001-639123/73.
P-PSDB; AAM51660.
Transgenic plants with improved heat stress tolerance, useful for
producing animal feed, oil and synthetic products -
Claim 4; Page -; 91pp; English.
The invention relates to a transgenic plant, comprising a genetic
construct comprising a promoter operatively linked to a nucleic acid
sequence (AA16057-AA16084) encoding a plant Heat Shock Protein (HSP)
family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
increased stress tolerance, especially to heat. The plant is a cereal,
grass, ornamental plant, crop plant, food plant, oil-producing plant, a
synthetic product-producing plant, an environmental waste absorbing
plant, an alcohol plant, a medicinal plant, a recreational plant and/or
an animal feed plant. In particular, the transgenic plant is cotton,
canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
Arabidopsis thaliana. The plants may be used to produce animal feed,
alcohol, crop, oil, medicine or a synthetic product.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained from GenBank using the Accession Number
reference provided in the specification.

Sequence 3942 BP; 834 A; 1111 C; 1236 G; 760 T; 1 other;

Alignment Scores:		
Pred. No.:	1.24e-261	3942
Score:	3657.00	770
Percent Similarity:	78.26%	Conservative: 76
Best Local Similarity:	71.23%	Mismatches: 60
Query Match:	80.37%	Indels: 177
DB:	22	Gaps: 11

US-09-812-350-17 (1-911) x AAI66074 (1-3942)

QY	1	MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu	20
Db	436	ATGAATCCGACCAACTTCACCCACAAGACGAACGAGCGCATCTGTGGGGCGCAGGATT	495
QY	21	AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer	40
Db	496	GCGGTGGAGCGCGCCACGCGCAGCTCACGCGCTGCACCTGCGCCGCGAGTGTGGCTGGCG	555
QY	41	AspProThrGlyIlePheProGlnAlaIleSerAla--GlyGlyGluAsnAlaAla	59
Db	556	GACAGGGCGGCAUCCCTCGCGCAGGCCATCACGGGGCGGTCTGGGGGGCGCAGGAGCGGCC	615
QY	60	---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro	78
Db	616	GCGGACTCGTTCGAGCGGTCTCAACAACCTCGCTCAAGAAGCTGCCGTGCGCAGTCCCCG	675
QY	79	ProProAspAspIleProAlaSerSerSerIleLysValIleArgAlaGlnAla	98
Db	676	CGCGCGGACTCGGTTTCGGCGCTCTACGGCGCTGATCAAGGTCTATCCGCGCGCGCGCAGTCC	735
QY	99	AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu	118
Db	736	GCGCAGAGAAGACGGCGGACTCGCACTCGCGTTCGACAGGTGTGTCTGCGCCCTGCTC	795
QY	119	GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys	138
Db	796	GAGGACTCGCAGATCTCCGACTGCCTCAAGAGCGCGCGGTGTCTCCGCGCGCGGTGGCG	855
QY	139	SerGluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAsp	158

956 GCCAGAGCTTCAGAAAGCTCCGCGCGGGAGGGCGCGCGCTGGAGTCCGCGTGGGGGAT 915
159 ThrAsnProGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeu 178
916 ACCAACTTCAGCGGCTCAAGACATACGCGCGGAGCTCTGTGAGCAGCGCGGAGCTT 975
179 AspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArg 198
976 GACCCCGTCATCGCGCGCAGCAGGAGATCCGCGCGTCTGTGCGCATTCCTCGCGCGC 1035
199 ThrLysAsnProValIleGlyGluProGlyValGlyLysThrAlaValValGlu 218
1036 ACCAAGAAATACCCCGCTCATCGCGAGCGCGCGTGGCAAGACGCGCTCGTGGAG 1095
219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
1096 GGCTTCGCGCAGCGCATCTTCGCGCGAGCTGCCAGTAACCTCTCGAGTCCGCTC 1155
239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGlu 258
1156 ATCGCGTTCAGCATGCGCGCTCTCAGGAGGTGGAGAGCGCGCAAGTACCGCGGAGTTCGAGGAG 1215
259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
1216 CGCTCAGCGCTGCTCAGGAGGTGGAGAGCGCGGAGGAGTCAATCTCTTCATC 1275
279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaLeu 298
1276 GACGAGATACACCTCTGCTGCGCGCGCGCAGCAGGAGGTTCATGACGCGCGCAAC 1335
299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
1336 CTGTTCAAGCCATGCTGCGAGGAGGACAGCTCAGGTGATCGCGCGCACCACTCGTGGAG 1395
319 GluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyr 338
1396 GAGTACCGCAAGTACGTGAGAGGAGCGAGCTTCAGCGCGGTTCAGCAGGTGTTTC 1455
339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyr 358
1456 GTCGCGAGCGGAGCTGCCGACACCGTCAGCATCTCAGCGGCTCAAGGAGAGTAC 1515
359 GluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaGlnLeuSer 378
1516 GAGGGCACATCGCGTGGAGATCCAGAGCGCGCTCTGTGTCGCGCACAGCTCTCC 1575
379 AlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
1576 GCGAGGTACATCATGCGGTGCGCACCTGCTGACAAAGCCATAGACTGTGTGACGAGGCC 1635
399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArg 418
1636 TGGCCCAATGTGAGGTGCGAGTTCGACAGCAGCGCGAAGAGATGATAACCTCGAGAGG 1695
419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSer 438
1696 AAGAAATCCAGTTCAGGTGAGTTCACGCGCTCGAAGAGGAGAGCAAGCCAGC 1755
439 LysAlaArgLeuIleGluVal----- 445
1756 AAGCCCGGTGATTGAGGT-GCGTGTCTCGAGCACTGAATTTCTCAAACAAGTCCTC 1814
445 ----- 445
1815 TTGCTGTGATGTTCTCTTTGAATTTTGATATACTAAGCTGCTTAACGTGCGCATTTGT 1874
446 -----ArgLysGluLeuAspLeuArgAspLysLeuGlnProLeuThrMetLysTyr 463
1875 CAGGTCAAGAGAAATGGACGATCTGAGGAGCAAGCTCGACCGCTGACCACTGAAGTAC 1934
464 ArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArgGluGluLeu 483
1935 AGGAAGAGAGAGAGAGAAATGTATGATCAGGAAGCTGAAGCAGCGCGCGAGGAGCTC 1994

QY 484 MetPheSerLeuGlnGluAlaGluArgTyrAspLeuAlaArgAlaAlaAspLeuArg 503
Db 1995 CAGTTTCACTTCGAGGAGCGCGAGCGCGATGAGCTTGGCCCTGTGGCGAGCTCAAG 2054
QY 504 TyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGlu 523
Db 2055 TACGGCGCTTCAGGAAATCGACGCTGCTATCTCAAGCTGGAG---AGCGAAACAGGG 2111
QY 524 GluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu----- 539
Db 2112 GAGAACCTGATGCTCCAGAAACCGTCGCGCTGCAACAAATTTGCAGAGGTATGTTATTAT 2171
QY 539 ----- 539
Db 2172 TCTTGTTCACCGTCACAAATAATTTGCAGAGCAAGTGCAGAAATTTGCGGATCGTCTCCTA 2231
QY 540 -----ValValSerAr 543
Db 2232 GTAGAGTAGTCGTGCGCGTCTGAAATGGTGTTCGTCTATGGCGAGGTGTGTAGCCG 2291
QY 543 GTrpThrGlyIleProValThrArgLeuGlyGlnAsnGlnLysGluArgLeuIleGlyLe 563
Db 2292 TTGACGCGGTATTCAGTGACCGCGCTTGGCCAGCAAGCAGAGAGGCTGTGGCT 2351
QY 563 uAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGl 583
Db 2352 GGCTGACAGGCTTCACAGAGGGTGTTCGCGCAGACAGAGGCTGTGAGCGCCCTCGCAGA 2411
QY 583 uAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLe 603
Db 2412 GCGGTGCTGAGTTCGAGGCGCGCTTTCGCGAGCCACAAACAGCCCACTGGCTCGTTCT 2471
QY 603 uPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLe 623
Db 2472 CTTCTGCTGCGTCCGACTGGCGTGGGAAAACCTGAGCTGGCCAGAGGCCCTAGCCCAACAGCT 2531
QY 623 uPheAspAspGluAsnLeuLeuValArgIleAspMetSerGluTyrMetGluGlnHisSe 643
Db 2532 GTTCGACGACGAGAACCTCTTGTTCGCGCATCGACATGTCGAGGTACATGGAGCAGCACTC 2591
QY 643 rValSerArgLeuIleGlyAlaProGly----- 653
Db 2592 GGTTCGCGCTCATCGAGCACCACCTGG-GTAAAGTAGCAGAAATGTCATGTCATCTTCG 2650
QY 653 ----- 653
Db 2651 TATTTTAACTGCGAATTGCGATGAATCTTGTTCGACAGTGGCGCGACCTGTGACAG 2710
QY 654 -TyrValGlyHisGluGluGlyGlnLeuThrGluAlaValArgArgArgProTyrCy 673
Db 2711 CTACGTGCGGCATGAAGAGGCTGGCGAGCTGACTGAACAAGTGGAGGAGGCGGTACAG 2770
QY 673 sValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuGlu 693
Db 2771 CGTGTCTGTTCACGAGTTCGAGAGAGGCCCTATGTTCGCGTGTTCACACACCTGTCTCA 2830
QY 693 nValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgTyrValAspPheArgAsnSe 713
Db 2831 GGTCTCTGACGACGCGAGGCTGACGATGGGCAAGGACGCGTGGACTTCAGGAACAC 2890
QY 713 rValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLy 733
Db 2891 CGTGATCATCATGACATCGAACTCGCGCGCGAGCACCTCTCGTGGGATGTGTGGCAA 2950
QY 733 sValThrMetGluValAlaArgAspCysValMet-Arg----- 745
Db 2951 GAATCTCATGAAGTGTGCTCGCATCTGGTTCATGAGGAGGTATGCAATTCGATAACGTT 3010
QY 745 ----- 745
Db 3011 CTGATCGTGTAGACATGTTCTCTTCGAGCGGTGATGCTCAATCGGATCGTTATTGCT 3070

746 -----GluValArgLysHisSheArgProGluLeuLeuAsnArgLeuAspGluLeuValV 764
 3071 GTGTGAGGTGAGGAGGCACTCCGCGCTGAGTCTGAACCGTCTGACGAGATCGTGA 3130
 764 alPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspV 784
 3131 TCTTCGATCTCTGTCCACGAGAGCTGAGAGGTGCTGCGCTTCAGATGAGGATG 3190
 784 alAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspT 804
 3191 TGGCGCTCGGTCTTCCGAAAGGGGCATCGTCTGCTGTGACCGACGCCCATGACCA 3250
 804 yrlleLeuAlaGluSerThrAspPro----- 812
 3251 TCATCTTGTCTCTCTTACGATCCGCTATGTGACCAATCCATGATTTGATCCATCTGAAT 3310
 812 ----- 812
 3311 TCGTCGGTGACCTGATGGTGTGACTCTCTTATCTTTCTTTGTGTGCTTCAACAACAG 3370
 813 ValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSer 832
 3371 GTGTATGGCGCGCGCAATCAGCAGGTGTGATCGAAGAGGTGTGACGAGCTCTCG 3430
 833 LysMetValValArgGluGluLeuAspGluAsnSerThrValTyrIleAspAlaGlyAla 852
 3431 AAGATGCTGTATCCAGGAGGAGATCGACGAGAACTGCACGGTCTTACATCGACGCCGCC 3490
 853 Gly-----AspLeuValTyrArgValGlu---SerGlyGlyLeuValAspAlaSerThr 869
 3491 GGCAAGGACGAGCTGTCTACAGGTGTGACCGGAGCGCGGTCTGTGTAACGCTGAGACG 3550
 870 GlyLysLysSerAspValLeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAla 889
 3551 GGGATGAAGTCGACATCTTGATCCAGTCCCAACAGCTCCACACGAGCGCTGCG 3610
 890 GlnAlaValLysLysMetArgIle---GluGluLeuAspAspAspAsnGluGlu 907
 3611 CAGCCCTCAAGAAGATGAGGATCATGAGGAGGACGAGGACCGCATGGAGGAG 3667

JUL 12

AAI66071

AAI66071 standard; DNA; 2045 BP.

AAI66071;

14-JAN-2002 (first entry)

Zea mays 101 kDa heat shock protein gene (GenBank: AF083327).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana; ds.

Zea mays.

W0200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-199116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

P-PSDB; AAM51670.

Transgenic plants with improved heat stress tolerance, useful for

PT producing animal feed, oil and synthetic products -
 XX Claim 4; Page -: 91pp; English.

XX The invention relates to a transgenic plant, comprising a genetic
 CC construct comprising a promoter operatively linked to a nucleic acid
 CC sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
 CC family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 CC increased stress tolerance, especially to heat. The plant is a cereal,
 CC grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 CC synthetic product-producing plant, an environmental waste absorbing
 CC plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 CC an animal feed plant. In particular, the transgenic plant is cotton,
 CC canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 CC Arabidopsis thaliana. The plants may be used to produce animal feed,
 CC alcohol, crop, oil, medicine or a synthetic product.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained from GenBank using the Accession Number
 CC reference provided in the specification.

SQ Sequence 2045 BP; 470 A; 532 C; 673 G; 370 T; 0 other;

Alignment Scores:

Pred. No.:	2,52e-176	Length:	2045
Score:	2502.50	Matches:	497
Percent Similarity:	93.65%	Conservative:	49
Best Local Similarity:	85.25%	Mismatches:	32
Query Match:	55.00%	Indels:	5
DB:	22	Gaps:	4

US-09-812-350-17 (1-911) x AAI66071 (1-2045)

QY 329 AlaPheGluArgArgPheGlnGlnValTyrValAlaGluProSerValProAspThrIle 348
 Db 3 GGTTCGAGCGCGGTTCAGCAGGTGTTCGCGGAGCGGAGCGGCGGATACCGTC 62
 QY 349 SerIleLeuArgGlyLeuLysGluLysTyrGluGlyHisHisGlyValArgIleGlnAsp 368
 Db 63 AGCATTTCTGAGGGGACTCAAGGAGAAGTACGAGGGGGACCATGGCGTGAGGATCCAGGAC 122
 QY 369 ArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArgTyrIleThrGlyArgHisLeuPro 388
 Db 123 CGCGCCCTCGTGTGTCGCGGCACAGCTCTCCGCGAGGTACATCATGGTGGCGCACCTGGCT 182
 QY 389 AspLysAlaIleAspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSer 408
 Db 183 GACAAAGCCATAGACCTGTTGGACGAGGCTTCGCGCAATGAGGGTGCAGCTCGACAGC 242
 QY 409 GlnProGluGluIleAspAsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHis 428
 Db 243 CAGCCGAGGAGATTGATAACTCTGGAGAGGAAGAGATCCAGCTTGAGTTCAGTCCAC 302
 QY 429 AlaLeuGluArgGluLysAspLysAlaSerLysAlaArgLeuIleGluValArgLysGlu 448
 Db 303 GCGCTCGAAGAGGAGAGACCAAGCCAGTAGTAACCCGCTGATTGAGGTTCAGGAAGGAA 362
 QY 449 LeuAspAspLeuArgAspLysLeuGlnProLeuThrMetLysTyrArgLysGluLysGlu 468
 Db 363 TTGGACGATCTGAGGACCAAGCTCGAGCCCTTGACCATGAAGTACAGGAGGAGAGAG 422
 QY 469 ArgIleAspGluIleArgArgLeuLysGlnLysArgGluGluLeuMetPheSerLeuGln 488
 Db 423 AGAATCGATGAGATCAGGAAGCTGAAGCAGCGCCGCGAGGAGTCCAGTTCACCTCCAG 482
 QY 489 GluAlaGluArgArgTyrAspLeuAlaArgAlaAlaAspLeuArgTyrGlyAlaIleGln 508
 Db 483 GAGCGGAGCGCGGATGACCTGCGCGGTGTGCGCGATCTCAAGTACGCGCGCTCCAG 542
 QY 509 GluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAenValMetLeu 528
 Db 543 GAATCGACGCGGATCTCCAGCTGAG---AGCGAAACACAGGGGAGACCTGATGCTC 599
 QY 529 ThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTrpThrGlyIlePro 548

600 ACCGAAACCGTCGCGCCCTGACAAATTCGAGAGGTGTCGAGCGGTATTTCCA 659
549 ValThrArgLeuGlyGlnAsnGluLeuGluArgLeuLeuLeuLeuLeuHis 568
660 GTGACCCGCGTGTGGCCAGACGACAGAGAGGTGTCGCGCCCTGCGTACAGGCTTCA 719
569 LysArgValValGlyGlnAsnGlnAlaValSerGluAlaLeuLeuLeuSer 588
720 CAGAGGTGTGTGGCCAGACGAGAGGTGTCGAGCGCGCTGCGAGGGCGTGTGAGTGC 779
589 ArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuProThr 608
780 AGGCGCGGTGTGGCCAGACGACAGAGGTGTCGAGCGCGCTGCTCTCTCTCTGCGACT 839
609 GlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspGluAsn 628
840 GCGGTGGGAAACTGAGCTGGCCAGGCGCTAGCCGACAGCTGTGTGACGAGAGAAC 899
629 LeuLeuValArgIleAspMetSerGlyThrMetGluGlnHisSerValSerArgLeuIle 648
900 CTTCTGTGCGCATCGACATGTCGAGTACATGAGCAGCAGCAGCTCGGTGCTCGCTCATC 959
649 GlyAlaProProGlyThrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArg 668
960 GGAGCACCATCTGCTGCTGCGCCATGAGAGGGTGGCAGCTGACTGAACAGGTGAGG 1019
669 ArgArgProThrCysValIleLeuPheAspGluValGluLysAlaHisValAlaValPhe 688
1020 AGGAGCGGTACAGCGTGTCTGTCAGAGGTGCGAGAGGCCCATGTGCGCGGTGTC 1079
689 AsnThrLeuLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrVal 708
1080 AACACCCGTGTCCAGGTCTCGACAGCGCAGGTGACGAGTGGCGAAGCAGCAGCGTG 1139
709 AspPheArgAsnSerValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAla 728
1140 GACTTCAGGAACACCGGTGATCATCATGACATGCAACCTCGCGCCGCGAGCACCTCTCGCT 1199
729 GlyLeuThrGlyLysValThrMetGluValAlaArgAspCysValMetArgGluValArg 748
1200 GGGATGTGGCGAAGACTCCATGAGGTGCTGCGATCTGTCATGCGAGGTGAGG 1259
749 LysHisPheArgProGluLeuLeuAsnArgLeuAspGluLeuValValPheAspProLeu 768
1260 AGGCACATTCGCGCCCTGAGTGTGCTGACCGCTCGACGAGATGCTGATCTCGATCTCTG 1319
769 SerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeu 788
1320 TCCACAGCAGCTGAGAGAGTGTGCTGCTTCAGATGAGGTGTCGCGGTGCTCTT 1379
789 AlaGluArgGlyValAlaLeuAlaValThrAspAlaLeuAspThrIleLeuAlaGlu 808
1380 GCGCAAGGGCATCGCTGCTGCTGTGACCGCGCATTCGACATCATCTGTCTCTC 1439
809 SerThrAspProValThrGlyAlaArgProIleArgArgThrMetGluLysValVal 828
1440 TCTTACATCCGCTGTACCGCGCGCGCCCAATAGAGGTGATCGAGAGAGGTGCTG 1499
829 ThrGluLeuSerLysMetValValArgGluGluLeuAspGluAsnSerThrValThrIle 848
1500 ACGAGTGTGAGAGTCTGTGATCCAGAGAGATCGACGAGACTGCGCGTCTACATC 1559
849 AspAlaGlyAlaGly-----AspLeuValThrArgValGlu---SerGlyGlyLeuVal 865
1560 GACCGCGCGCGGTGAGAGAGTGTCTACAGGTGCGACCGGAGCGCGTCTGCTG 1619
866 AspAlaSerThrGlyLysLysSerAspValLeuLeuHisIleAlaAsnGlyProLysArg 885
1620 AAGCTGAGAGGGGATGAGTGTGAGTCTGATCCAGTCTCCACCGCTCCACGAG 1679
886 SerAspAlaAlaGlnAlaValLysLysMetArgIle---GluGluLeuGluAspAsp 904

Db 1680 AGCGACGCTGCGCGCCGTCAAGAGATGAGATCATGAGGAGGACGAGCGCATG 1739
QY 905 AsnGluGlu 907
Db 1740 GACGAGGAG 1748
RESULT 13
AAI66072
ID AAI66072 standard; DNA; 2045 BP.
XX AAI66072;
AC AAI66072;
XX 14-JAN-2002 (first entry)
XX Zea mays heat shock protein 101 gene (GenBank: AF077337).
XX Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
KW Arabidopsis thaliana; da.
XX Zea mays.
XX WO200170929-A2.
XX 27-SEP-2001.
XX 20-MAR-2001; 2001WO-088836.
XX 20-MAR-2000; 2000US-190769P.
XX 18-APR-2000; 2000US-198116P.
XX (ARCH-) ARCH DEV CORP.
XX Lindquist S, Queitsch C, Vierling B;
XX WPI; 2001-639123/73.
XX P-PSDB; AAM51663.
XX Transgenic plants with improved heat stress tolerance, useful for
producing animal feed, oil and synthetic products -
XX Claim 4; Page -; 91pp; English.

CC The invention relates to a transgenic plant, comprising a genetic
construct comprising a promoter operatively linked to a nucleic acid
sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
increased stress tolerance, especially to heat. The plant is a cereal, a
grass, ornamental plant, crop plant, food plant, oil-producing plant, a
synthetic product-producing plant, an environmental waste absorbing
plant, an alcohol plant, a medicinal plant, a recreational plant and/or
an animal feed plant. In particular, the transgenic plant is cotton,
canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
Arabidopsis thaliana. The plants may be used to produce animal feed,
alcohol, crop, oil, medicine or a synthetic product.
CC Note: the sequence data for this patent did not form part of the printed
specification, but was obtained from GenBank using the Accession Number
reference provided in the specification.

SQ Sequence 2045 BP; 470 A; 532 C; 673 G; 370 T; 0 other;

Alignment Scores:

Pred. No.:	2,52e-176	Length:	2045
Score:	2502.50	Matches:	497
Percent Similarity:	93.65%	Conservative:	49
Best Local Similarity:	85.25%	Mismatches:	32
Query Match:	55.00%	Indels:	5
DB:	22	Gaps:	4

US-09-812-350-17 (1-911) x AAI66072 (1-2045)

QY 329 AlaPheGluArgArgPheGlnGlnValTyValAlaGluProSerValProAspThrIle 348
|||||

3 GCGTTCCAGCGCGGTTCCAGCAGGTTGTCGCGGAGCGCGGTCGCCGATACCGTTC 62
 349 SerIleLeuArgGlyLeuLysGluValThrGluGlyHisHisGlyValAlaGluLeuAsp 368
 63 AGCATTCGAGGGGACTCAGAGAGAGTACAGGGGACCATGGCGTGAGGATCCAGGAC 122
 369 ArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArgTyrIleThrGlyArgHisLeuPro 388
 123 CGCGCCCTCGTGTGCGCGCAGCATCTCCGCGAGGTACATCATGGGTGCGCACCTGCT 182
 389 AspLysAlaIleAspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSer 408
 183 GACAAGCCATAGACCTCGTGGAGCAGGCTCGGCCAATGTGAGGGTGCAGCTCGACAGC 242
 409 GlnProGluGluIleAspAsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHis 428
 243 CAGCCGAGGAGATTGATTAACCTGGAGAGAGAGAGATCCAGCTTGAGGTCCAGCTCCAC 302
 429 AlaLeuGluArgGluLysAspLysAlaSerLysAlaArgLeuIleGluValArgLysGlu 448
 303 GCGCTCGAGAGAGAGAGAGCAAGGCCAGTAAAGCCCGGCTGATTGAGTCCAGGAAGAA 362
 449 LeuAspLeuArgGluAspLysLeuGlnProLeuThrMetLysTyrArgLysGluLysGlu 468
 363 TTGGACGATCTGAGGAGCAAGCTGAGCCCTGAGCCATGAAGTACAGGAAGGAGAGAG 422
 469 ArgIleAspGluIleArgArgLysGlnLysArgGluGluLeuMetPheSerLeuGln 488
 423 AGAATCGATGAGTCAGAGAGCTGAGAGAGCGCGCGAGAGCTCCAGTTCACCTGCAG 482
 489 GluAlaGluArgArgTyrAspLeuAlaArgAlaAlaAspLeuArgTyrGlyAlaIleGln 508
 483 GAGCGCAGCGCGGATGACCTGCGCCGCTGTGGCCGATCTCAAGTACGCGCGCTCCAG 542
 509 GluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAsnValMetLeu 528
 543 GAAATCGAGCGCGGATCTCAAGCTTGAG---AGCGAAACAGGGGAGAACTGATGCTC 599
 529 ThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThrGlyIlePro 548
 600 ACCGAAACCGTCCGCGCTGACAAATTCAGAGGTGGTGGACCGGTGGACGGGTATCCA 659
 549 ValThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHis 568
 660 GTGACCCGCTTGCGCCAGAACCAACAGAGAGGCTGGTGGCGCTGGCTGACAGGCTTCA 719
 569 LysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSer 588
 720 CAGAGGTGGTGGCGCAGACAGAGCTGTGAGCGCGCTCGCAGAGCGCGTGTGAGGTGG 779
 589 ArgAlaGlyLeuGlyArgAlaGlnProThrGlySerPheLeuPheLeuGlyProThr 608
 780 AGGCGCGTCTTGCGAGCCCAACAGCCCACTGGCTGTTCTCTTCTTCTGGTCCGACT 839
 609 GlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsn 628
 840 GCGGTGGGGAACCTGAGCTGGCCCAAGGCCCTAGCCGAACAGCTGTTCCAGCAGCAGAAC 899
 629 LeuLeuValArgIleAspMetSerGluTyrMetGluGlnHisSerValSerArgLeuIle 648
 900 CTTCTGTGCGCATCGATGCGAGTACATGGAGCAGCAGCTCGGTGCTGCGCTCATC 959
 649 GlyAlaProProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArg 668
 960 GGAGCACCACTGGTACGTCGCGCATGAAGAGGGTGGCAGCTGACTGAACAAGTCAGG 1019
 669 ArgArgProTyrCysValIleLeuPheAspGluValGluLysAlaHisValAlaValPhe 688
 1020 AGGAGGCGGTACAGCGTATCTGTGTCAGAGGTTCGAGAGGCGCCATGTGCGCGTGTTC 1079
 689 AsnThrLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrVal 708
 1080 AACACCTGCTCCAGTCTCCAGCAGCGAGGCTGACGGATGGGCAAGGAGGAGCGGTG 1139

QY 709 AspPheArgAsnSerValIleIleMetThrSerAsnLeuGluValAlaGluHisLeuLeuAla 728
 DB 1140 GACTTCAGGAACACCGTGATCATCATGATCATCATCGAAGCTGGCGCCGAGCACCTCTCGT 1199
 QY 729 GlyLeuThrGlyLysValThrMetGluValAlaAlaArgAspCysValMetArgGluValArg 748
 DB 1200 GGGATGTTGGGCAAGAACTCCATGAAGTTCGCTCGCATCTGGTCTATGAGGAGGTGAG 1259
 QY 749 LysHisPheArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspProLeu 768
 DB 1260 AGCATCTTCGCGCTGAGCTGCTGAACCGTCTCGACGAGATCGTATCTTCGATCTCTG 1319
 QY 769 SerHisAspGlnLeuArgLysValAlaAlaArgLeuGlnMetLysAspValAlaValArgLeu 788
 DB 1320 TCCACGAGCAGCTGAGGAGGTGCTGCTCCCTTCAGATGAGGATGTGGCCGCTCTCTT 1379
 QY 789 AlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGlu 808
 DB 1380 GCCGAAAGGGGATCTCTGCTGTGACCGCAGCGCATTCGACATCATCTTGTCTCTC 1439
 QY 809 SerTyrAspProValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValVal 828
 DB 1440 TCTTACGATCCGCTGTACGCGCGCGCCCAATAAGAGGTGGATCGAGAAGAGGGTGGT 1499
 QY 829 ThrGluLeuSerLysMetValValArgGluGluIleAspGluAsnSerThrValTyrIle 848
 DB 1500 ACCAGCTGTGAGAGTGTGATCCAGGAGGAGATCGACGAGACTGCGACGCTTACATC 1559
 QY 849 AspAlaGlyAlaGly-----AspLeuValTyrArgValGlu---SerGlyGlyLeuVal 865
 DB 1560 GACGCGCGCGCGGTAAGGACGAACCTGGTCTACAGGTGCGACCGGAGCGCGTCTGGTG 1619
 QY 866 AspAlaSerThrGlyLysSerAspValLeuIleHisIleAlaAsnGlyProLysArg 885
 DB 1620 AACGCTGAGACGGGGATGAAGTCGACATCTCGATCCAGTCCCCACAGCTCCACGAG 1679
 QY 886 SerAspAlaAlaGlnAlaValLysLysMetArgIle---GluGluIleGluAspAsp 904
 DB 1680 AGCAGCGTCCGCGAGCCCTCAAGAAGATGAGGATCATGAGGAGGAGGAGCGGCGCATG 1739
 QY 905 AsnGluGlu 907
 DB 1740 GACGAGGAG 1748
 RESULT 14
 AAI66083
 ID AAI66083 standard; DNA; 7726 BP.
 AC AAI66083;
 XX
 XX 14-JAN-2002 (first entry)
 DT
 XX
 DE Leishmania donovani heat shock protein 100 gene (GenBank: Z94053).
 XX Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 KW Arabidopsis thaliana; ds.
 XX
 OS Leishmania donovani.
 XX
 FN WO200170929-A2.
 PD 27-SEP-2001.
 XX
 PF 20-MAR-2001; 2001WO-US08836.
 XX
 PR 20-MAR-2000; 2000US-190769P.
 PR 18-APR-2000; 2000US-198116P.
 XX
 PA (ARCH-) ARCH DEV CORP.
 XX
 PI Lindquist S, Queitsch C, Vierling E;

4060 GTTGTGCTGGAGCATATCCCGTGCACAAAGCTGACCGACCGACCGAGCCG 4119
 560 LeuIleGlyLeuAlaLeuArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsn 579
 4120 CTGCTGCACCTGGCCGACCACTGCACTCGGTGTAAGGGCCAGGATGAGCGGTGAGC 4179
 580 AlaValSerGluAlaLeuArgSerArgAlaGlyLeuGlyGlnGlnGlnProThr 599
 4180 CGTGTGCGGAGGCTATCTGCGCTCAGTGGCGGCTGCGCCGCTGCGACCGGCCACC 4239
 600 GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeu 619
 4240 GGCTCTTCTGCTTCTCGGCCCCACCGCGGTGGCAAGACAGAGCTGTCCAGGCGCTC 4299
 620 AlaGluGlnLeuPheAspGluAsnLeuLeuValArgLysLeuValArgLysLeuMet 639
 4300 GCTCAAGAGCTCTTCGACGATGCAAGATGATGCTGCGACTGACATGACGAGTACATG 4359
 640 GluGlnHisSerValSerArgLeuLeuGlyValAlaProProGlyTyrValGlyHisGluGlu 659
 4360 GAGCAGACCTCTGTGGCGGCTGATCGCGGCCCGCCCGCGGTGCTGATGACGAGGAA 4419
 660 GlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGlu 679
 4420 GGTGGCCAGCTGACGAGCGGTGGCGCTGCGCCATACACGCTGCTGCTGATGAG 4479
 680 ValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspGlyArg 699
 4480 GTGGAGAGCGCCACCGACGCTTCAAGCTACTGCTGACGATCATCATGATCATCC 4539
 700 LeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleLeuMetThrSer 719
 4540 TTGACTGACTCGACGCGCGCAACCGTGAATTTCTGCTACGATCATCATGATCATCC 4599
 720 AsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAla 739
 4600 AACTGGTGGCGAGTATCTGCAGAACATGACCTCACCGAAGCGCTACGAGTGGCA 4659
 740 ArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeu 759
 4660 CAGACACAGGTGATGGCGAGGTGAGGAAGTTCTTTCGCCCGAGTTCATCAACCGACTG 4719
 760 AspGluLeuValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeu 779
 4720 GATGACATCATCTCTCCGCTCTCTGGGCTTAAAGGAGATGACCGGATCATCGACCTC 4779
 780 GlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAsp 799
 4780 ATCACCAGGAGGCTCTATGGCCGCTCTCAAGGACAGTCCATCCGCGGTTCCTCAGAA 4839
 800 AlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyLysArgProIle 819
 4840 GAGGCTAAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4899
 820 ArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGlu 839
 4900 CGCGCGTGGGTTGAGAGAACATCACAACAGAGCTCAGCGGATGATCATCTTCGCAAGAG 4959
 840 IleAspGluAsnSerThrValTyrIle 848
 4960 CTGTGCGCGCAACAGCAGCGGTGAAGTG 4986

SU15
 166084

AAI66084 standard; DNA; 5997 BP.

AAI66084;

14-JAN-2002 (first entry)

Leishmania major heat shock protein Hsp100 gene (GenBank: Z38058).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana; da.
 Leishmania major.
 WO200170929-A2.
 27-SEP-2001.
 20-MAR-2001; 2001WO-US08836.
 20-MAR-2000; 2000US-190769P.
 18-APR-2000; 2000US-198116P.
 (ARCH-) ARCH DEV CORP.
 Lindquist S, Queitsch C, Vierling E;
 WPI: 2001-639123/73.
 P-PSDB; AAM51668.
 Transgenic plants with improved heat stress tolerance, useful for
 producing animal feed, oil and synthetic products -
 Claim 4; Page -; 91pp; English.
 The invention relates to a transgenic plant, comprising a genetic
 construct comprising a promoter operatively linked to a nucleic acid
 sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
 family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 increased stress tolerance, especially to heat. The plant is a cereal,
 grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 synthetic product-producing plant, an environmental waste absorbing
 plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 an animal feed plant. In particular, the transgenic plant is cotton,
 canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 Arabidopsis thaliana. The plants may be used to produce animal feed,
 alcohol, crop, oil, medicine or a synthetic product.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained from GenBank using the Accession Number
 reference provided in the specification.
 SQ Sequence 5997 BP; 1182 A; 1830 C; 1645 G; 1340 T; 0 other;
 Alignment Scores:
 Pred. No.: 6,46e-160 Length: 5997
 Score: 2289.00 Matches: 461
 Percent Similarity: 72.08% Conservative: 151
 Best Local Similarity: 54.30% Mismatches: 227
 Query Match: 50.31% Indels: 10
 DB: 22 Gaps: 5
 US-09-812-350-17 (1-911) x AAI66084 (1-5997)
 Qy 3 ProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAlaVal 22
 Db 2187 CCAGAA---TGAAGCAGGCGCGCGCTCGGACTGTGATGCTCGCATGCGGCGACCTGGC 2243
 Qy 23 AsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSerAspPro 42
 Db 2244 AAGAAGGCAACCGCTACCTCGACCCCGTGCATCTCGCTACGTCATGTTTGAAGATGAA 2303
 Qy 43 ThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGlnSerAla 62
 Db 2304 AACAGTCTTCTCCCGTGGCGCGCAAACTCGGT-----GCCGCTCAGTC 2351
 Qy 63 GluArgValIleAsnGlnAlaLeuLysLysLeuSerGlnSerProProAspAsp 82
 Db 2352 AAGCAGGACTGGAGGCGCGCTGTCGACGCGATTTCCACACGATGCTGCGCCACGCGAG 2411
 Qy 83 IleProLysSerSerLeuIleLysValIleLysValIleArgArgAlaGlnAlaGlnLysSer 102
 Db 83 IleProLysSerSerLeuIleLysValIleLysValIleArgArgAlaGlnAlaGlnLysSer 102

2412 CGCGGCCCCAACTCGGACATGATCGGTCTCATGAACGCGGAGCAGGAGCGTGTGCC 2471
103 ArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAspSerGln 122
2472 CTCGGTGACACCTCATCGCGCTGACCACTTCTCTGGCCCTGCGAGAGCAAGGAG 2531
123 IleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGluValGlu 142
2532 GTTGAAGAATCTGGATGCTGCGAGGGCTGGAAGAAGGCGATTGCGCGCCAGCTGCTC 2591
143 LysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsnPheGln 162
2592 GAGATGCC-----AAGGGAAAGAAAGATACCTCCGACTTCCAAAGCAGCAACTACGAG 2645
163 AlaLeuLysThrThrGlyArgAspLeuValGluGlnAla-----GlyLysLeuAspPro 180
2646 TCGCTGAACAGTAGCAGGATGATCTCTGCAAGCAGGCGGAGGAGGAAAGCTGAGCCCG 2705
181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgThrLys 200
2706 GTTATTGTCGCGCAGACGAGATTCTGCGCACTATCCGCTGCTGTCACGCCGACCAAG 2765
201 AsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeu 220
2766 AACAAACCGGTGCTTATCGGGAGCCCTGGAGTGGGTAAAGACCGCGATTGAGAGGCGCAT 2825
221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
2826 CGCGACAGGTGGTGGCGAGCGAGCTGCCGACACCTCTCCGGTATTCCGACTTCTCA 2885
241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysThrArgGlyGluPheGluGluArgLeu 260
2886 CTGACATGGCGCGCTGATCGCGGTGCGCAAGTACCGCGCGAGTTCGAGGAGCGCTG 2945
261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
2946 AAGCCCTGTGTAATGAGGTGAGGAAGCGCAATAAGATCAATCTCTTATTGTCGAG 3005
281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300
3006 ATCCACCTGCTGTCGCGCGCGCAAGTCCGCGCGCGATGACGCGCGGATTCGCTG 3065
301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGlyTyr 320
3066 AAGCCGTGTGTCGCGCGTGGCGAGCTGCGCAGCATCGGTGCCCAACCGTTCGAGGAGTAC 3125
321 ArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnValTyrValAla 340
3126 CGCCAGTACCTGGAGAGGACGCGCTTCGAGCGCGCTTCATGCCAGTGCAGGTGCAT 3185
341 GluProSerValProAspThrIleSerIleLeuArgLysGluLysTyrGluGly 360
3186 GAGCCGTCACTGAGGAGTGCAGCATCTCGTGGTCTGAAGACCGCTACGAGCAG 3245
361 HisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaGlnLeuSerAlaArg 380
3246 CACCACGTGTGAGATCAGGACAGCGCGTGTGTCGCGCGCGAGTTCGTCGCGCGC 3305
381 TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla 400
3306 TACATTACGAACCGCTTCCTGCGCGGATAGCGGATGACCTCATGATGAGCGTGC 3365
401 AsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArg 420
3366 AAGTGGCGCTGACGTGTGTCGCGCGCGCGGAAATCGACGCCCTCGAGCGCAAGAG 3425
421 MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla 440
3426 CGCCAGTGGAGATCGAGGAGAGGGCAGCTGACGCGGCAAGGACCATCGGCCAAGGAG 3485
441 ArgLeuIleGluValArgLysGluLeuAspLeuArgAspLeuAspLysAlaProLeuThr 460
3486 CGGCTGAAGCCCTGAAGCGGAGATTCAGAGGTGAGGAGAACTTGGCCCTCTACTT 3545

461 MetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArg 480
3546 GCCAAGTACGAACGAGGCGTGGCGGTATTGAGAGCTGCGCAACGCGCAAAAGCTG 3605
481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAlaAla 500
3606 GACGAGAAGAAGGTGAGCTGAGCGCGGAGCGAGCGGATCGCGCATGGAGACGCGACA 3665
501 AspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu---GluGly 519
3666 GACCTCAATACCGGTTCATCCCAATCTCCAGGACAGATCCGATCCCTCAAGGAGCG 3725
520 ThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu 539
3726 ATCGAGACGACGAGCGGAGCATGCTGCGAGGGCACCGTGACCGGACGAGCATCCGACC 3785
540 ValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArg 559
3786 GTTGTGTCGCGCTGACCAATATCCCGTGCACAAAGCTGAGCCAGCAGCAACGCGAGCGC 3845
560 LeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsn 579
3846 CTGTTGCACTTCCGCGCAGCATGTCCTCGTGTGAAGGCGCAGGATGAGCGGTGAGC 3905
580 AlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThr 599
3906 CGTGTGCGGAGCGCATCTGCGCTCAGCTGCGGCTGGCCGCTTGGAGCCGCGCCACC 3965
600 GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeu 619
3966 GGCCTTTCTCTGTTCTCGTCCACCGCGCTGGGCAAGCAGAGCTATCCAAAGCGCGTT 4025
620 AlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluTyrMet 639
4026 GCCTCAGAGCTCTTCGACGACGCCAAGTACATGTTGCGACTTGACATGAGCGATACATG 4085
640 GluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHisGluGlu 659
4086 GAGCAGCACTCTGTGGCGCGCTGATCGGACCCCGCGGGGTACGTGCGCCACGAAGA 4145
660 GlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGlu 679
4146 GGTGCCAGCTGACGAGCGCGTGGCGCTGCCATACACGCTGGTGGTGTCTGATGAA 4205
680 ValGluLysAlaHisValAlaValPheAsnThrLeuGlnValLeuAspAspGlyArg 699
4206 GTGGAAGAAGCGCACCCGACGCTCTCAACGTACTGTGCGAGGTGCTCGACGCGGCGG 4265
700 LeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSer 719
4266 TTGACTGACTCGACGCGCGCAACCGTGATTTCTGCAACAGCATCATCATGACATCC 4325
720 AsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAla 739
4326 AACTGGGTGGCAGTATCTGCAGAACATGCAGACCTTCACGGAAGCGGTACGAAGTGGCA 4385
740 ArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeu 759
4386 CAGGACAGGTGATGGCGGAGGTAAAGAAAGTTCTTTCGCGCGGAGTTTCATCAACCGACTG 4445
760 AspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeu 779
4446 GACGACATCATCTCTTCCTCGCTCTCTGGGCTTAAAGAGATGACCGGCATCATCGACCTC 4505
780 GlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAsp 799
4506 ATCAGAGAGGTCAATGGCGCGCTCAGGACAGTCCATCGGGGTCTCCCTCAGCGAA 4565
800 AlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIle 819
4566 GAGCCCAAGCAGTACGTATTGGAGTTCGCGCTTCGACGCGGATATGGGCGCTCGCCGCTA 4625

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 12, 2004, 22:27:44 ; Search time 72 Seconds
(without alignments)
2008.332 Million cell updates/sec

US-09-812-350-17

fect score: 4550

quence:

1 MNPEKFTKHTNETATHEL.....VKMKRIEIEDDNEEMTED 911

ring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

1107863 seqs, 158726573 residues

al number of hits satisfying chosen parameters: 1107863

imum DB seq length: 0

imum DB seq length: 2000000000

t-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

abase : A_Geneseq_13Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneexp-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneexp-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneexp-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneexp-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneexp-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneexp-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneexp-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneexp-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneexp-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneexp-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneexp-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneexp-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneexp-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneexp-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneexp-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneexp-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneexp-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneexp-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneexp-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneexp-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneexp-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneexp-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneexp-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneexp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ult No.	Score	Query Match	Length DB	ID	Description
1	4550	100.0	911	22	AA151659 Arabidopsis thalia
2	4550	100.0	911	22	AA151660 Triticum aestivum
3	4545	99.9	911	22	ABP81189 Arabidopsis thalia
4	4003	88.0	911	22	AA151662 Glycine max heat sh
5	3896	85.6	909	22	AA151665 Nicotiana tobacum
6	3883.5	85.4	912	22	AA151663 Zea mays heat choc
7	3883.5	85.4	912	22	AA151669 Zea mays heat choc
8	3822.5	84.0	913	22	AA151664 Triticum aestivum
9	3790	83.3	918	22	AA151661 Triticum aestivum

10	2502.5	55.0	582	22	AA151670 Zea mays 101 kDa h
11	2431	53.4	668	22	AA151666 Arabidopsis thalia
12	2298	50.5	869	22	AA151667 Leishmania donovan
13	2289	50.3	867	22	AA151668 Leishmania major 1
14	2288	50.3	872	23	ABG77915 High level promote
15	2236.5	49.2	866	23	ABG49337 Listeria monocytog
16	2224.5	48.9	868	22	AA151658 Trypanosoma brucei
17	2224	48.9	883	22	AA151651 Synchococcus CLPB
18	2204	48.4	861	22	AA151651 Salmonella typhi c
19	2194	48.2	857	20	AA151651 Staphylococcus aur
20	2193	48.2	857	22	AA151651 E. coli cellular p
21	2185	48.0	889	23	ABP66002 Bifidobacterium lo
22	2162.5	47.5	856	22	AA151651 Haemophilus influe
23	2158.5	47.4	848	23	ABJ04699 Mycobacterium tube
24	2145	47.1	852	22	AA151656 Corynebacterium gl
25	2145	47.1	852	22	AA151656 C glutamicum prote
26	2141.5	47.1	869	22	AA151656 Staphylococcus aur
27	2141.5	47.1	882	22	AA151656 Staphylococcus aur
28	2140.5	47.0	885	22	AA151656 Propionibacterium
29	2133	46.9	854	22	AA151656 Pseudomonas aerugi
30	2113	46.4	859	20	AA151656 Amino acid sequenc
31	2108	46.3	871	23	ABP39134 Staphylococcus epi
32	2107.5	46.3	859	24	ABP79777 N. gonorrhoeae ami
33	2099.5	46.1	867	23	ABP54868 Lactococcus lactis
34	2033.5	44.7	977	22	AA151671 Phaseolus lunatus
35	1993.5	43.8	870	23	AA151671 Fungal heat resist
36	1938	42.6	867	22	AA151671 C. trachomatis CT1
37	1931	42.4	867	20	AA151671 Protein involved i
38	1912	42.0	866	22	AA151671 C. pneumoniae CT1
39	1912	42.0	872	20	AA151671 C. pneumoniae prot
40	1890	41.5	908	22	AA151657 Saccharomyces cere
41	1884.5	41.4	856	22	AA151657 Helicobacter pylor
42	1876.5	41.2	856	22	AA151657 Helicobacter pylor
43	1839	40.4	908	14	ABP33429 E. coli heat shock
44	1830	40.2	823	23	ABP39236 Staphylococcus epi
45	1828	40.2	817	22	AA151678 S. epidermidis ope

ALIGNMENTS

RESULT 1	AA151659	AA151659 standard; Protein; 911 AA.
ID	AA151659	Arabidopsis thaliana heat shock protein 101 (GenBank: P42730).
AC	AA151659;	Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
XX		canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
XX		Arabidopsis thaliana.
DT	14-JAN-2002	(first entry)
XX		Arabidopsis thaliana.
DE		Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
XX		canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
KW		Arabidopsis thaliana.
XX		Arabidopsis thaliana.
OS		Arabidopsis thaliana.
XX		Arabidopsis thaliana.
PN	WO200170929-A2.	
XX		
PD	27-SEP-2001.	
XX		
PF	20-MAR-2001; 2001WO-US08836.	
XX		
PR	20-MAR-2000; 2000US-190769P.	
XX		
PA	18-APR-2000; 2000US-198116P.	
XX		
XX	(ARCH-) ARCH DEV CORP.	
PI	Lindquist S, Queitsch C, Vierling E;	
XX	WPI; 2001-639123/73.	
DR	N-PSDB; AA166065, AA166066.	
XX		

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 2; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AA166057-AA166084). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 911 AA;

Query Match 100.0%; Score 4550; DB 22; Length 911;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 911; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNPEFTHKTNETIATAHELVNAGHAQTPHLALISDPTGIFPOAISAGGENAAQ 60
1 MNPEFTHKTNETIATAHELVNAGHAQTPHLALISDPTGIFPOAISAGGENAAQ 60
61 SAERVINOALKLPSPQPPDDIPASSLSIKVIRRAQAQKSRGTHLAVQILMGLLED 120
61 SAERVINOALKLPSPQPPDDIPASSLSIKVIRRAQAQKSRGTHLAVQILMGLLED 120
121 SOIRDLNLEVGATARVSEKELRGKGGKVESASGDTNFQALKTGYGRDLVEQAGKLD 180
121 SOIRDLNLEVGATARVSEKELRGKGGKVESASGDTNFQALKTGYGRDLVEQAGKLD 180
181 VTGRDEIRRVRIISRTKNNPVLIPGGVGTAVVEGLAQRIVKGVDPNSLTDVRLIS 240
181 VTGRDEIRRVRIISRTKNNPVLIPGGVGTAVVEGLAQRIVKGVDPNSLTDVRLIS 240
241 LDMGALVAGAKYRGEFFERLKSVEDEDAEKGVLFTIDEIHLVLGAGKTEGSDAANLF 300
241 LDMGALVAGAKYRGEFFERLKSVEDEDAEKGVLFTIDEIHLVLGAGKTEGSDAANLF 300
301 KEMLARGOLRCIGATTLEBYRKYVEKDAFERFQOVVVAEPSPDITSIIRGLKEKYE 360
301 KEMLARGOLRCIGATTLEBYRKYVEKDAFERFQOVVVAEPSPDITSIIRGLKEKYE 360
361 HGVRIQDRALINAAQLSARYITGRHLPDKAIDLVEACANVRVOLDQPEIDNLERKR 420
361 HGVRIQDRALINAAQLSARYITGRHLPDKAIDLVEACANVRVOLDQPEIDNLERKR 420
421 MQLTELHALEREKDKASKARLIEVRKELDLDRDKLQPLTWKYRKEKERIDEIRLKKQR 480
421 MQLTELHALEREKDKASKARLIEVRKELDLDRDKLQPLTWKYRKEKERIDEIRLKKQR 480
481 BELMPSLOAERYDLARAADLYRGAIOEVESAIQLEGTSSEENVMLTENVPHEIAEV 540
481 BELMPSLOAERYDLARAADLYRGAIOEVESAIQLEGTSSEENVMLTENVPHEIAEV 540
541 VSRWTGIPVTRIGQNEKERLIGLDRHLKRVVQGNVAVSEALIRSRAGLGRQQPTG 600
541 VSRWTGIPVTRIGQNEKERLIGLDRHLKRVVQGNVAVSEALIRSRAGLGRQQPTG 600
601 SFLFLGPTGVGKTELAKALAEQLFDENLLVRIDMSYMEQHSVRLIGAPPGVYGHGEG 660
601 SFLFLGPTGVGKTELAKALAEQLFDENLLVRIDMSYMEQHSVRLIGAPPGVYGHGEG 660
661 GOLTEAVRRPVCVILFDEVEKARHVAVENTLLQVLDGRLTDGQRTVDFRNSVLIIMTSN 720

Db 661 GOLTEAVRRPVCVILFDEVEKARHVAVENTLLQVLDGRLTDGQRTVDFRNSVLIIMTSN 720
Qy 721 LGAHLLAGLTGKVTMEVARDCVMEVRKHFRELLNRLDEIVVDFPLSHDQLRKVRLQ 780
Db 721 LGAHLLAGLTGKVTMEVARDCVMEVRKHFRELLNRLDEIVVDFPLSHDQLRKVRLQ 780
Qy 781 MKDVAVLAERGVVALAVTDAALDYILAESVDPVYGARPIRRWMEKKVTVLSKVVVEEII 840
Db 781 MKDVAVLAERGVVALAVTDAALDYILAESVDPVYGARPIRRWMEKKVTVLSKVVVEEII 840
Qy 841 DENSTVYIDAGDLVYRVESGGLVDASTGKSDVLIHTANGPKRSDAAQAVKMKRIEII 900
Db 841 DENSTVYIDAGDLVYRVESGGLVDASTGKSDVLIHTANGPKRSDAAQAVKMKRIEII 900
Qy 901 EDDNNEEMIED 911
Db 901 EDDNNEEMIED 911

RESULT 2

AA166057
ID AA166057 standard; Protein; 911 AA.

AA166084

14-JAN-2002 (first entry)

Triticum aestivum heat shock protein 101 (GenBank: AAD22629).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
Arabidopsis thaliana.

Triticum aestivum.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08936.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

N-PSDB; AA166074.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 2; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AA166057-AA166084). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 911 AA;

Query Match	100.0%;	Score 4550;	DB 22;	Length 911;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 911;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
1	MNPEKFTHTKTETATATAHELAVNACHAQTPLHLAGALISDPTGIFFOAISSAGGENAAQ	60			
1	MNPEKFTHTKTETATATAHELAVNACHAQTPLHLAGALISDPTGIFFOAISSAGGENAAQ	60			
61	SAERVINQALKKLPSSQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLIMGLLED	120			
61	SAERVINQALKKLPSSQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLIMGLLED	120			
121	SOIRDLNNEVGAVATARKVSEVEKLRGKGGKVESASGDTNFQALKTGYGRDLVEQAGKLDP	180			
121	SOIRDLNNEVGAVATARKVSEVEKLRGKGGKVESASGDTNFQALKTGYGRDLVEQAGKLDP	180			
181	VIGRDEEIRRVVRLISRRTKNNPVLIGPGVGKTAVVEGLAQRIVKGDVNSLTDVRLIS	240			
181	VIGRDEEIRRVVRLISRRTKNNPVLIGPGVGKTAVVEGLAQRIVKGDVNSLTDVRLIS	240			
241	LDMGALVAGAKYRGGEFEERLKSVLKEVEDAEGKVLFIIDEHLVLGAGKTEGSDMAANLF	300			
241	LDMGALVAGAKYRGGEFEERLKSVLKEVEDAEGKVLFIIDEHLVLGAGKTEGSDMAANLF	300			
301	KPMLARGQLRCIGATTILSEYKYVEKDAAPERRFQOVYVAEPSVPTDISILRGLKEKYE	360			
301	KPMLARGQLRCIGATTILSEYKYVEKDAAPERRFQOVYVAEPSVPTDISILRGLKEKYE	360			
361	HGVRIQDRALINAQLSARVITGRHLPDKAIDLVDEACANVRVQLDSQPEEDINLERK	420			
361	HGVRIQDRALINAQLSARVITGRHLPDKAIDLVDEACANVRVQLDSQPEEDINLERK	420			
421	MLBIELHALEREKDKASKARLIEVRKELDDRLKLOPLTMKYRKEKERIDEIRLKQKR	480			
421	MLBIELHALEREKDKASKARLIEVRKELDDRLKLOPLTMKYRKEKERIDEIRLKQKR	480			
481	EELMFSQAEARRVDLARAADLRVCAQVESAIQAQEGTSSEENVMLTENVGPEHIAEV	540			
481	EELMFSQAEARRVDLARAADLRVCAQVESAIQAQEGTSSEENVMLTENVGPEHIAEV	540			
541	VSRWTGIPVTRLGQNEKERLIGLADRLHKRVGQONAVNAVSAIRLRSRAGLGRAQOPTG	600			
541	VSRWTGIPVTRLGQNEKERLIGLADRLHKRVGQONAVNAVSAIRLRSRAGLGRAQOPTG	600			
601	SFLPLGPTGVKTELAKALAEQLPDDENLLVRLIDMSEYMEQHSVSRILGAPPGVGHGEG	660			
601	SFLPLGPTGVKTELAKALAEQLPDDENLLVRLIDMSEYMEQHSVSRILGAPPGVGHGEG	660			
661	GOLTEAVRRPYCVILFDEVEKAHVAFTNLLQVLDGRLTDGGRTPFRNSVITMNS	720			
661	GOLTEAVRRPYCVILFDEVEKAHVAFTNLLQVLDGRLTDGGRTPFRNSVITMNS	720			
721	LGAEHLLAGLTGKVTMEVARDVCMREVRKFRPELLNRLDEIVVEDPLSHDQLRKVARLQ	780			
721	LGAEHLLAGLTGKVTMEVARDVCMREVRKFRPELLNRLDEIVVEDPLSHDQLRKVARLQ	780			
781	MKDVAVRLAERGVAVLAVTDAAALDYVILAESYDPVYGARPIRRMWEKKVVTLSQVVRREI	840			
781	MKDVAVRLAERGVAVLAVTDAAALDYVILAESYDPVYGARPIRRMWEKKVVTLSQVVRREI	840			
841	DENSTVYIDAGAGLVYRVESGGLVDASTGKSDVLIHANGPKRSAAQAVKQKRIEII	900			
841	DENSTVYIDAGAGLVYRVESGGLVDASTGKSDVLIHANGPKRSAAQAVKQKRIEII	900			
901	EDDDNEEMIED 911				
901	EDDDNEEMIED 911				

ABP81189 standard; Protein; 911 AA.

XX	ABP81189;				
XX	AC				
XX	27-FEB-2003	(first entry)			
XX	Arabidopsis thaliana protein #17 modulated by PTGS.				
XX	Posttranscriptional gene silencing; PTGS; plant; transformation.				
XX	Arabidopsis thaliana.				
XX	W0200281695-A2.				
XX	17-OCT-2002.				
XX	05-APR-2002; 2002WO-EP03806.				
XX	06-APR-2001; 2001US-282049P.				
XX	(SYGN) SYNGENTA PARTICIPATIONS AG.				
XX	(FRIE-) FRIEDRICH MIESCHER INST.				
XX	Zhu T, Glazov EA, Meins F, Wang X, Chang H;				
XX	WPI; 2003-103337/09.				
XX	N-PSDB; AB242033.				
XX	Novel polynucleic acid segment useful for modulating gene expression				
XX	within a cell by posttranscriptional gene silencing, and for augmenting				
XX	a plant cell genome -				
XX	Claim 53; Page 186-190; 438pp; English.				
XX	The invention relates to a novel isolated polynucleic acid segment				
XX	modulated within a cell by posttranscriptional gene silencing (PTGS). The				
XX	invention specifically relates to a method to identify an expression				
XX	product that is modulated by PTGS. The polynucleotide is useful for				
XX	modulating the gene expression within a cell by PTGS, by introducing the				
XX	polynucleic acid into a cell and expressing the nucleic acid segment in				
XX	the cell to form a product. The polynucleic acid segment is also useful				
XX	for augmenting a cell genome, and for augmenting a plant genome, by				
XX	contacting a plant cell with the segment to produce a transformed plant				
XX	cell, and growing the transformed plant cell to produce a differentiated				
XX	transformed plant. The sequences shown in ABP81173 - ABP81298 represent				
XX	the product of a segment of A. thaliana cDNA modulated by PTGS.				
XX	Sequence 911 AA;				
XX	Query Match 99.9%; Score 4545; DB 24; Length 911;				
XX	Best Local Similarity 99.9%; Pred. No. 0;				
XX	Matches 910; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
QY	1 MNPEKFTHTKTETATATAHELAVNACHAQTPLHLAGALISDPTGIFFOAISSAGGENAAQ	60			
Db	1 MNPEKFTHTKTETATATAHELAVNACHAQTPLHLAGALISDPTGIFFOAISSAGGENAAQ	60			
QY	61 SAERVINQALKKLPSSQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLIMGLLED	120			
Db	61 SAERVINQALKKLPSSQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLIMGLLED	120			
QY	121 SOIRDLNNEVGAVATARKVSEVEKLRGKGGKVESASGDTNFQALKTGYGRDLVEQAGKLDP	180			
Db	121 SOIRDLNNEVGAVATARKVSEVEKLRGKGGKVESASGDTNFQALKTGYGRDLVEQAGKLDP	180			
QY	181 VIGRDEEIRRVVRLISRRTKNNPVLIGPGVGKTAVVEGLAQRIVKGDVNSLTDVRLIS	240			
Db	181 VIGRDEEIRRVVRLISRRTKNNPVLIGPGVGKTAVVEGLAQRIVKGDVNSLTDVRLIS	240			
QY	241 LDMGALVAGAKYRGGEFEERLKSVLKEVEDAEGKVLFIIDEHLVLGAGKTEGSDMAANLF	300			
Db	241 LDMGALVAGAKYRGGEFEERLKSVLKEVEDAEGKVLFIIDEHLVLGAGKTEGSDMAANLF	300			
QY	301 KPMLARGQLRCIGATTILSEYKYVEKDAAPERRFQOVYVAEPSVPTDISILRGLKEKYE	360			

301 KPMALGQRCIGATTLEEKYKVEKDAFERFOQVVAEPSPDTTISILRGLKEKYEG 360
 361 HHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEIDNLERK 420
 361 HHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEIDNLERK 420
 421 MQLTEELHALEREKDKASARLIEVREKELDDLRDKLOPLTMKYRKEKERIDIRLKKQR 480
 421 MQLTEELHALEREKDKASARLIEVREKELDDLRDKLOPLTMKYRKEKERIDIRLKKQR 480
 481 EELMFLQBAERYDLARAADLYGAIQVESALAEQLEGTSEENVMLTENVGPHIAEV 540
 481 EELMFLQBAERYDLARAADLYGAIQVESALAEQLEGTSEENVMLTENVGPHIAEV 540
 541 VSRWTGIPVTRLGQNEKERLIGLADRLHRRVVGQNAVNAVSEAILRSGRAGRAQQPTG 600
 541 VSRWTGIPVTRLGQNEKERLIGLADRLHRRVVGQNAVNAVSEAILRSGRAGRAQQPTG 600
 601 SFLFLGPTGVGKTELAKALAEQIFDDENLLVRIDMSEYMEQHSVSRLLICAPPGYVGHGEG 660
 601 SFLFLGPTGVGKTELAKALAEQIFDDENLLVRIDMSEYMEQHSVSRLLICAPPGYVGHGEG 660
 661 GQTEAVRRRPPYCVILFDEVEKAHVAVFNTLLQVLDGRLTDGQGRTVDRNSVLIIMTSN 720
 661 GQTEAVRRRPPYCVILFDEVEKAHVAVFNTLLQVLDGRLTDGQGRTVDRNSVLIIMTSN 720
 721 LGAEHLLAGLTKVMTMEVARDCMREVRKHFPELNLRLDEIVVDFPLSHDQLRVARLQ 780
 721 LGAEHLLAGLTKVMTMEVARDCMREVRKHFPELNLRLDEIVVDFPLSHDQLRVARLQ 780
 781 MKDVAVRLAERGVAVLAVTDAALDYILAESYDPVYGARPIRRWMEKKVTELSKVVREIEI 840
 781 MKDVAVRLAERGVAVLAVTDAALDYILAESYDPVYGARPIRRWMEKKVTELSKVVREIEI 840
 841 DENSTVYIDAGADLVYRVESGGLVDASTGKSDVLIHANGPKESDAAQVKKWRIEIEI 900
 841 DENSTVYIDAGADLVYRVESGGLVDASTGKSDVLIHANGPKESDAAQVKKWRIEIEI 900
 901 EDDNNEEMIED 911
 901 EDDNNEEMIED 911

JLT 4
 AAM51662

AAM51662 standard; Protein; 911 AA.

AAM51662;

14-JAN-2002 (first entry)

Glycine max heat shock protein (GenBank: AAM56338).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana.

Glycine max.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

DR WPI; 2001-639123/73.
 DR N-PSDB; AAI66067.
 PT Transgenic plants with improved heat stress tolerance, useful for
 PT producing animal feed, oil and synthetic products -
 XX Claim 2; Page -; 91pp; English.
 PS The invention relates to a transgenic plant, comprising a genetic
 CC construct comprising a promoter operatively linked to a nucleic acid
 CC sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP)
 CC family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 CC increased stress tolerance, especially to heat. The plant is a cereal,
 CC grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 CC synthetic product-producing plant, an environmental waste absorbing
 CC plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 CC an animal feed plant. In particular, the transgenic plant is cotton,
 CC canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 CC Arabidopsis thaliana. The plants may be used to produce animal feed,
 CC alcohol, crop, oil, medicine or a synthetic product.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained from GenBank using the Accession Number
 CC reference provided in the specification.
 XX Sequence 911 AA;

Query Match 88.0%; Score 4003; DB 22; Length 911;
 Best Local Similarity 87.0%; Pred. No. 1.5e-289;
 Matches 794; Conservative 72; Mismatches 39; Indels 8; Gaps 6;

QY 1 MNPEKFTHTKNTETATAHNAVAGHAQFTPLHAGALISDPTGIFPQAISAA-GGENAA 59
 DB 1 MNPEKFTHTKNTETATAHNAVAGHAQFTPLHAGALISDPTGIFPQAISAA-GGENAA 59
 QY 60 QSAERVINQAKLPGQSPDPDIPASSSLIKVIRAAQAQKSRGTHLAVDOLIMGLE 119
 DB 61 RAVERNINQAKLPGQSPDPDIPASSSLIKVIRAAQAQKSRGTHLAVDOLIMGLE 120
 QY 120 DSQIRDLNMEVGAVATARKVSEKLGKGGKVESASGDTNFQALKTYGRDVEQAGKLD 179
 DB 121 DSQIGDLLKEAGVAVAKVSEVDKLGKGGKVESASGDTNFQALKTYGRDVEQAGKLD 180
 QY 180 PVIGRDEIRRVRIILSRRTKNNPVILIGEPGVGKTAVVEGLAQRIIVKGDVPNSLTDVRLI 239
 DB 181 PVIGRDEIRRVRIILSRRTKNNPVILIGEPGVGKTAVVEGLAQRIIVKGDVPNSLTDVRLI 240
 QY 240 SLDMGALVAGAKYRGEFEERLKSILKEVEDAGKVLIFDEIHLVLGAGKTGSGMDAANL 299
 DB 241 ALDMGALVAGAKYRGEFEERLKSILKEVEDAGKVLIFDEIHLVLGAGKTGSGMDAANL 300
 QY 300 FKPMALARGQLRCIGATTLEERYKYVEKDAAFERRFOQVVAEPSPDTTISILRGLKEKYE 359
 DB 301 FKPMALARGQLRCIGATTLEERYKYVEKDAAFERRFOQVVAEPSPDTTISILRGLKEKYE 360
 QY 360 GHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEIDNLERK 419
 DB 361 GHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDSQPEIDNLERK 420
 QY 420 RMQLELELHALEREKDKASARLIEVREKELDDLRDKLOPLTMKYRKEKERIDIRLKKOK 479
 DB 421 RMQLELELHALEREKDKASARLIEVREKELDDLRDKLOPLTMKYRKEKERIDIRLKKOK 480
 QY 480 REELMPSLOEABERRYDLARAADLYGAIQVESALAEQLEGTSEENVMLTENVGPHIAE 539
 DB 481 REELMPSLOEABERRYDLARAADLYGAIQVESALAEQLEGTSEENVMLTENVGPHIAE 539
 QY 540 VVSRWTGIPVTRLGQNEKERLIGLADRLHRRVVGQNAVNAVSEAILRSGRAGRAQQPT 599
 DB 540 VVSRWTGIPVTRLGQNEKERLIGLADRLHRRVVGQNAVNAVSEAILRSGRAGRAQQPT 599
 QY 600 GSFLFLGPTGVGKTELAKALAEQIFDDENLLVRIDMSEYMEQHSVSRLLICAPPGYVGHGEE 659
 DB 600 GSFLFLGPTGVGKTELAKALAEQIFDDENLLVRIDMSEYMEQHSVSRLLICAPPGYVGHGEE 659

Note: The sequence data for this patent did not form part of the printed publication, therefore this patent document is not a prior art reference to the extent that it describes any such sequence data. This patent document is not a prior art reference to the extent that it describes any such sequence data.

RESULT 6

AA51663

AA51663 standard; Protein; 912 AA.

AA51663;

14-JAN-2002 (first entry)

Zea mays heat shock protein 101 (GenBank: AAD25223).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana.

Zea mays.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

N-PSDB; AA166072.

Transgenic plants with improved heat stress tolerance, useful for
 producing animal feed, oil and synthetic products -

Claim 2; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic
 construct comprising a promoter operatively linked to a nucleic acid
 sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP)
 family amino acid sequence (AA51651-AA51671). The transgenic plant has
 increased stress tolerance, especially to heat. The plant is a cereal,
 grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 synthetic product-producing plant, an environmental waste absorbing
 plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 an animal feed plant. In particular, the transgenic plant is cotton,
 canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 Arabidopsis thaliana. The plants may be used to produce animal feed,
 alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained from GenBank using the Accession Number
 reference provided in the specification.

Sequence 912 AA;

Query Match 85.4%; Score 3883.5; DB 22; Length 912;

Best Local Similarity 84.6%; Pred. No. 1.3e-280;

Matches 772; Conservative 76; Mismatches 58; Indels 7; Gaps 6;

1 MNPKFTKNTETATTAHELAVNAGHAQFTPLHLAGALISDPTGIPQAISSA-GGENAA 59

1 MNPDNFTKNTETATTAHELAVNAGHAQFTPLHLAGALISDPTGIPQAISSA-GGENAA 60

60 -QSAERVINOALKLPSPQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLIMGLL 118

61 GDSFERNLNNLKLPLSPQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLIMGLL 120

119 EDSQIDLLNEGVATARYKSEVKLRGKKGKVSASGDTNFQALKTYGRDLVDAQKL 178

121 EDSQISDCLKEAGVSAARVRAELKRGGEGRVESASGDTNFQALKTYGRDLVDAQKL 180

179 DPVIGDEIRVRVILSRRTKNNPVLIGEPGVGTAVVEGLAQRIVRGDPVSNLLDVRL 238

Db 181 DPVIGDEIRVRVILSRRTKNNPVLIGEPGVGTAVVEGLAQRIVRGDPVSNLLDVRL 240
 Qy 239 ISLDGALVAGAKYRGFEERLKSVLKEVEDAEKGVILFIDEIHLVLGAKTEGSMDAAN 298
 Db 241 IALDMLGALVAGAKYRGFEERLKAFLKEVEAEKGVILFIDEIHLVLGAKTEGSMDAAN 300
 Qy 299 LFKPMLARGQLRCIGATTILBEYRKYVEKDAAFERFQOVVAVPSVPTTISILRGLKEKY 358
 Db 301 LFKPMLARGQLRCIGATTILBEYRKYVEKDAAFERFQOVVAVPSVPTTISILRGLKEKY 360
 Qy 359 EGHGVRIQDRALVAAQALSARYITGSHLPDKAIDLVDDEACANVRVQLDSQPEIDNLER 418
 Db 361 EGHGVRIQDRALVAAQALSARYITGSHLPDKAIDLVDDEACANVRVQLDSQPEIDNLER 420
 Qy 419 KRMOLFELEHALREKDKASKARLIEVRKELDLRLDKLOPLTMKYRKEKERIDEIRLQK 478
 Db 421 KRIQLEVELEHALREKDKASKARLIEVRKELDLRLDKLOPLTMKYRKEKERIDEIRLQK 480
 Qy 479 KREELMFSLQEAERRYDLARAADIRYGAIOEVESAIQALEGTSSEENVMLTENVGPEHIA 538
 Db 481 RREELQFTLQEAERRYDLARAADIRYGAIOEVESAIQALEGTSSEENVMLTENVGPEHIA 539
 Qy 539 EVSRWTGIPVTRLGQNEKERLIGLADRLHKRVVQNOAVNAVSEAILRSRAGLGRAPQ 598
 Db 540 EVSRWTGIPVTRLGQNEKERLIGLADRLHKRVVQNOAVNAVSEAILRSRAGLGRAPQ 599
 Qy 599 TGSFLFLGPTGVGKTELAKALAEQLPDENLLVRIDMSEYMEQHSVSRLLGAPGVGHE 658
 Db 600 TGSFLFLGPTGVGKTELAKALAEQLPDENLLVRIDMSEYMEQHSVSRLLGAPGVGHE 659
 Qy 659 EGGQLEAVRRPYCVILFDEVEKAHVAVENTLLQVLDDGRLTDGQRTVDFRNSVIMT 718
 Db 660 EGGQLEAVRRPYCVILFDEVEKAHVAVENTLLQVLDDGRLTDGQRTVDFRNSVIMT 719
 Qy 719 SNLGAELHLAGLTGKVTMEVARDCVREVRKRPPELLNRLDEIVDFPLSHDQLKVAR 778
 Db 720 SNLGAELHLAGLTGKVTMEVARDCVREVRKRPPELLNRLDEIVDFPLSHDQLKVAR 779
 Qy 779 LQMKDVAVRLAERGVALLVDAALDYTLAESYDPVYGARPIRREMEKKVVTLSKMWVRE 838
 Db 780 LQMKDVAVRLAERGVALLVDAALDYTLAESYDPVYGARPIRREMEKKVVTLSKMWVRE 839
 Qy 839 EIDENSTVYIDAGAG--DLVYRVE--SGGLVDASTGKSDVLIHIANGPKRSDAQAQVKM 895
 Db 840 EIDENSTVYIDAGAG--DLVYRVE--SGGLVDASTGKSDVLIHIANGPKRSDAQAQVKM 899
 Qy 896 RI-REIEDDDNEE 907
 Db 900 RIMEEDEDGMDDEE 912

RESULT 7
 AA51669
 ID AA51669 standard; Protein; 912 AA.
 XX
 AC AA51669;
 XX
 DT 14-JAN-2002 (first entry)
 XX
 DE Zea mays heat shock protein HSP101 (GenBank: AAD33606).
 XX
 KW Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 KW Arabidopsis thaliana.
 XX
 OS Zea mays.
 XX
 FN WO200170929-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 20-MAR-2001; 2001WO-US08836.
 XX

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AA16057-AA16084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AA51651-AA51671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing

plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 913 AA;

Query Match 84.0%; Score 3822.5; DB 22; Length 913;

Best Local Similarity 83.3%; Pred. No. 4.5e-276;

Mismatches 761; Conservative 75; Mismatches 69; Indels 9; Gaps 6;

1 MNPDKTHKTNETIATAHELAVNAGHAQTPHLHAGALISDPTGIPQAISSAGGEN--A 58
 1 MNPDKTHKTNETIATAHELAVNAGHAQTPHLHAGALISDPTGIPQAISSAGGENA 60
 59 AQAERVINQALKKLPSPQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLL 118
 61 AESFERVAGALKRPLSPQSPDDTPVASTSLKAVRAQAQKSRGDSHLAVDQIMGLL 120
 119 EDSQIDILNEVCVATARKSEVKRGKGGKVSASGDTNFOALKTYGRDLVBOAGKL 178
 121 EDQISDALKEAGISAARKAEYKELGGDNRRVESASGDTNFOALKTYGRDLVBOAGKL 180
 179 DPVIGDEIRRVVRLISRTKNNPVLIGPGVGTAVVEGLAQRIVKGDVNSITDVL 238
 181 DPVIGDEIRRVVRLISRTKNNPVLIGPGVGTAVVEGLAQRIVKGDVNSITDVL 240
 239 ISLDMGALVAGKRGFEERLKSILKVEDAEGKVLFDIHLVLGAGTKTEGSDAAN 298
 241 VALDMGALVAGKRGFEERLKSILKVEDAEGKVLFDIHLVLGAGTKTEGSDAAN 300
 299 LKPMELARGQLRCIGATTLEEVKVEKDAAFERRFQVYVAEPSVPTISLRLGK 358
 301 LKPMELARGQLRCIGATTLEEVKVEKDAAFERRFQVYVAEPSVPTISLRLGK 360
 359 EGHGVRIQDRALINAAQLSARYITGRHLDPKADILVDEACANVRVOLDSPEDNLER 418
 361 EGHGVRIQDRALINAAQLSARYITGRHLDPKADILVDEACANVRVOLDSPEDNLER 420
 419 KMOLEIEHLERKOKKASKARLIEVRKELDDRLKPLTMKYRKEKERIDEIRLQ 478
 421 KRIQLEVLHLEKOKKASKARLIEVRKELDDRLKPLTMKYRKEKERIDEIRLQ 480
 479 KEELMFSLOAERRYDLARAADLRYGATQEVESATQLEGTSSEENVMTENVPHEIA 538
 481 REELQFTLOAERRMDLARVADLRYGALQEVDAALQLEGTG-ENLMLETVTGPDQIA 539
 539 EYVSRWTGIPVTRLGONEKERLIGLADRLHKRVGQNOAVNAVSEAILRSRAGLRQOP 598
 540 EYVSRWTGIPVTRLGONEKERLIGLADRLHKRVGQNOAVNAVSEAILRSRAGLRQOP 599
 599 TGSFLFLGTGKTELAKALAEQFDENLIVRIMSEYMEQHSVSLIGAPPYVGH 658
 600 TGSFLFLGTGKTELAKALAEQFDENLIVRIMSEYMEQHSVSLIGAPPYVGH 659
 659 EGQQLTEAVRRPYPYVILLFDEVEKHAFAVNTLLQVDDGRLTDQGRVDFRNSVIIMT 718
 660 EGQQLTEAVRRPYPYVILLFDEVEKHAFAVNTLLQVDDGRLTDQGRVDFRNSVIIMT 719
 719 SNLGAHLLAGTGTGKTEVARDVCMREVRKFRPELLNRLDEIVVDFPLSHDQLRKVAR 778
 720 SNLGAHLLAGTGTGKTEVARDVCMREVRKFRPELLNRLDEIVVDFPLSHDQLRKVAR 778
 779 LOMKQVAVRLAERGVALAVTDALDIYIAESVDYPYGARPIRRMWEKVVTELKVVRE 838
 779 LOMKQVAVRLAERGVALAVTDALDIYIAESVDYPYGARPIRRMWEKVVTELKVVRE 838
 839 EIDENSTVYIDAGAG--DLAVRVES--GGLVDASTGKSDVLIHLIANGPKRSDAAQAVK 895

Db 839 EIDENSTVYIDAAAPSKDELTVGVKHGVLNARTGHKSDILIQVSPGAVGGDAHAHVKKM 898

QY 896 RI--BEIEDDNEE 907

Db 899 KIMQDSGVBDDME 912

RESULT 9

AAM51661

ID AAM51661 standard; Protein; 918 AA.

XX AC AAM51661;

XX 14-JAN-2002 (first entry)

XX Triticum aestivum 101kDa heat shock protein (GenBank: AAC83689).

XX Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;

XX canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;

XX Arabidopsis thaliana.

XX Triticum aestivum.

XX WO200170929-A2.

XX 27-SEP-2001.

XX 20-MAR-2001; 2001WO-US08836.

XX 20-MAR-2000; 2000US-190769P.

XX 18-APR-2000; 2000US-198116P.

XX (ARCH-) ARCH DEV CORP.

XX Lindquist S, Queitsch C, Vierling E;

XX WPI; 2001-639123/73.

XX N-PSDB; AAI66075.

XX Transgenic plants with improved heat stress tolerance, useful for

XX producing animal feed, oil and synthetic products -

XX Claim 2; Page -; 91pp; English.

XX The invention relates to a transgenic plant, comprising a genetic

XX construct comprising a promoter operatively linked to a nucleic acid

XX sequence (AAI66075-AAI66084) encoding a plant Heat Shock Protein (HSP)

XX family amino acid sequence (AAM51651-AAM51671). The transgenic plant has

XX increased stress tolerance, especially to heat. The plant is a cereal,

XX grass, ornamental plant, crop plant, food plant, oil-producing plant, a

XX synthetic product-producing plant, an environmental waste absorbing,

XX plant, an alcohol plant, a medicinal plant, a recreational plant and/or

XX an animal feed plant. In particular, the transgenic plant is cotton,

XX canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or

XX Arabidopsis thaliana. The plants may be used to produce animal feed,

XX alcohol, crop, oil, medicine or a synthetic product.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained from GenBank using the Accession Number

XX reference provided in the specification.

XX SQ Sequence 918 AA;

Query Match 83.3%; Score 3790; DB 22; Length 918;

Best Local Similarity 81.7%; Pred. No. 1.2e-273;

Mismatches 752; Conservative 86; Mismatches 70; Indels 12; Gaps 7;

QY 1 MNPDKTHKTNETIATAHELAVNAGHAQTPHLHAGALISDPTGIPQAISSAGGENA-- 59

Db 1 MNPDKTHKTNETIATAHELAVNAGHAQTPHLHAGALISDPTGIPQAISSAGGENA 60

QY 60 -QSAERVINQALKKLPSPQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLL 118

Db 61 GDSFERVLGALRLKPLSPQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLL 120

[illegible]

Zea mays.

420 RHEPELNLDELDEIVIPDPLSHEQLRKVARLQMKDVAVRLAERGIALAVTDAALDIILSL 479
 809 SYDPVYGARIPRRWMEKKVVTLSKVVREIDENSTVYIDAGAG--DLVYRYE-SGGILV 865
 480 SYDPVYGARIPRKWKIEKRVVTLQSKMLQIEIDENCTVYIDAAAGKDELIVYRVDSSGGLV 539
 866 DASTGKKSDVLIHANGPKRSDAAQAVKQRI--EEIEDDDNNEE 907
 540 NAETGKMSDILIQVPTSTSRSDAAQAVKWKIMEEDEDGMDDEE 582

RESULT 11

AAM51666
 AAM51666 standard; Protein; 668 AA.

AAM51666;

14-JAN-2002 (first entry)

Arabidopsis thaliana heat shock protein like (GenBank: CAB46061).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana.

Arabidopsis thaliana.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.
 18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

Transgenic plants with improved heat stress tolerance, useful for
 producing animal feed, oil and synthetic products -

Claim 2; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic
 construct comprising a promoter operatively linked to a nucleic acid
 sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP)
 family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
 increased stress tolerance, especially to heat. The plant is a cereal,
 grass, ornamental plant, crop plant, food plant, oil-producing plant, a
 synthetic product-producing plant, an environmental waste absorbing
 plant, an alcohol plant, a medicinal plant, a recreational plant and/or
 an animal feed plant. In particular, the transgenic plant is cotton,
 canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
 Arabidopsis thaliana. The plants may be used to produce animal feed,
 alcohol, crop, oil, medicine or a synthetic product.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained from GenBank using the Accession Number
 reference provided in the specification.

Sequence 668 AA;

Query Match 53.4%; Score 2431; DB 22; Length 668;
 est Local Similarity 70.8%; Pred. No. 1.9e-172;
 atches 499; Conservative 76; Mismatches 90; Indels 40; Gaps 6;

1 MNPEKTHKTNETIATAHELAVNAGHAQFTPLHLAGALISDPTGIFFOAISSAG-GENAA 59
 1 MNDLKFDPNVKLILASARSHAMLSHGQVTEPLHLGVTLSLTSVFFYRAITSAGDGISA 60

QY 60 QSAERVINOALKKLPSQSPPPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQIMGLLE 119
 Db 61 QSVNVINQSLFKL-----TKRNLGDTKVGVAVLVLSLE 95
 QY 120 DSQIRDLILNEGVGATARVKSEVKLRGKGGKKBVSASGDTNFQALKTYGRDLVEQAGKLD 179
 Db 96 DSQISDLVLEAGVVPKVKSEVKLR-----GEVILRALKTYGTDLVEQAGKLD 144
 QY 180 PVIGRDEIRVRVILSRRTKNPNVLIGEPGVGKTAVVEGLAQRIVRKGDVNSLTDVRLI 239
 Db 145 PVIGRHRIRRVIEVLSRRTKNPNVLIGEPGVGKTAVVEGLAQRIKGLDVPINLTGVKLI 204
 QY 240 SLDMGALVAGAKYGEFEERLKSVLKEVEDAEGKVILFIDEIHLVLGAGKTEGSMDAANL 299
 Db 205 SLEFGAMVAGTTLRGQFEERLKSVLKAVEBAQGVILFIDEIHMALGACKASGSDAAKL 264
 QY 300 FKPMIARGQLRCIGATTLLEEVRYKVKVEDAAFERRFQOVVVAEVPDITISILGLKEKYE 359
 Db 265 LKPMIARGQLRFIGATTLLEEVRYTHVEKDAFERRFQOVVVAEVPDITISILGLKEKYE 324
 QY 360 GHGVRIQDRALINAAQLSARYITGRHLDPKALDLVDEACANVRVQLDSQPEEIDNLERK 419
 Db 325 GHGVRIQDRALVLSAQLSERVITGRRLDPKALDLVDESAHVKAQLDIQPEEIDSLEK 384
 QY 420 RMQLEIEIHALEREK-DKASKARLIEVRKELDDLRLDKLOPLTMKYRKEKERIDEIRLQK 478
 Db 385 VMQLEIEIHALEREKDKASEARLSEVRKELDDLRLDKLEPLTTIKYKKEKKIINETRLQK 444
 QY 479 KRELMFSLQEAERRVYDLARAADLYGAIQEVESAIAQLEGTSSEENVMLTENVGPEHIA 538
 Db 445 NRDDLMTALQEAERQHDVPKAAVLYKGAIQEVESAIAKLE-KSAKDNVMLTEIVGPEHIA 503
 QY 539 EVYSRWTCIIVTRLGQNEKERLIGLADRLHKRVVGQNAVNAVSEAILRSRAGLRAQOP 598
 Db 504 EVYSRWTCIIVTRLGQNEKERLISLADKLHERVVGQDEAVKAVAAAILRSRVLGRPOOP 563
 QY 599 TGSFLFLGPTGVGKTELAKALAEQLFDENLIVRIDMSEYMEOHVSYSRLIGAPPG-VYGH 657
 Db 564 SGSFLFLGPTGVGKTELAKALAEQLFDSENLLVRLDMSEYNDKFSVNLKIGAPPGYIGH 623
 QY 658 EGGQLTEAVRRRPPYCVILFDEVEKARVAVFNTLLQVLDGRLTD 702
 Db 624 EGGQLTEPVRPPYCVILFDEVEKTHVTVFNTLLQVLDGRLTD 668

RESULT 12

AAM51667
 ID AAM51667 standard; Protein; 869 AA.

AC AAM51667;

14-JAN-2002 (first entry)

Leishmania donovani heat shock protein 100 (GenBank: CAB08073).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
 canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
 Arabidopsis thaliana.

Leishmania donovani.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

N-PSDB; AAI66083.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 2; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAM51651-AAM51671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or plant, an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 869 AA;

Query Match 50.5%; Score 2298; DB 22; Length 869;
Best Local Similarity 54.4%; Pred. No. 2.3e-162;
Matches 462; Conservative 153; Mismatches 224; Indels 10; Gaps 5;
3 PEKTHKNTETATAHNAVNAHQPTPLHAGALISDPTGIPPOAISSAGGENAQA 62
6 PE-WTQASDLMAARTALKKANGYLDVHLAVVPEDENSLASRVVKLG----AASV 60
63 ERVINQALKLPSPSPDDIPASSSLIKVIRRAQAQKRGDTHLAVDQIMGLSDSQ 122
61 KDGLEARVDIPTQMPAPTQPRNSDMWRVMTAEQERVALGDTLMAADHFLALHESKE 120
123 IRDLNVEGVATARKVEKLRCKEKGKVESASGDTNFOALKTYGRLVQEA--GKLDP 180
121 VGRILDAAGAKKAIRATILEMR--KGGKINSFQDDNYESLNKYAVDLCKQAEKGKLDP 178
181 VIGRDEIRVRVRLSRRTKNNPVLIGEPGVGKTA VVEGLAQRIVKGDVPSNLTVDRLIS 240
179 VIGRADILRTIRVLSRRTKNNPVLIGEPGVGKTA VVEGLAQQVVRGDPDILSGIRFS 238
241 LDMGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDRHLVLAGKTEGSDAANLF 300
239 LDMGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDRHLVLAGKSDGDAANLL 298
301 KPMIARGQLRCIGATTILEYRKVYKVDAAFERFQVYVAERPSVPTISILRGLKEKYG 360
299 KPIIARGELATIGATTILEYRKVYKVDAAFERFQVYVAERPSVPTISILRGLKDRYSQ 358
361 HHGVRQDRALINAQLSARYITGRHLPKDAIDLVDCAENRVQDSQPEEDINLERK 420
359 HHGVRQDRALINAQLSARYITGRHLPKDAIDLVDCAENRVQDSQPEEDINLERK 418
421 MQLEIEHLALEREKDKASKARLLEVRKELDDLDKLOPLTMKYRKEKEIDEIRLKQR 480
419 RQLEIEKALQDKDSAKERKAVKAEIQKVEKGLPLLAKYQERGRIDELQATQAKL 478
481 EELMFSIQAEERYDLARADLYRGAIQEVESAIAQL--EGTSSEENVMLTENVEPHAE 539
479 DEKKVKLRAERMRDMETADLKNVPILODRIRSLKEEIKQKATMLHGTVTETDIAT 538
540 VVSRWTGIPVTRIGQNEKERLIGLADRLHKKRVVQGNQAVNSEAILSRAGIGRAQOPT 599
539 VVSRWTNIPVTKLSQTERELLHDLQHLRVKQDEAVSRVAEILSRAGLARSDEPT 598
600 GSFLFGTGVGKTELAKALAEQLFDENLLVRIDMSYMBQHSVSRSLTGAPPVGVGHEE 659

Db 599 GSFLFGTGVGKTELAKALAEQLFDENLLVRIDMSYMBQHSVSRSLTGAPPVGVGHEE 658
Qy 660 GQQLTEARRRPPYCVILFDEVERKAHVAVFNTLLQVLDGRLTDGQGRITVDFRNSVLIIMTS 719
Db 659 GQQLTEPVRRRPYTVVLLDEVERKAHVAVFNTLLQVLDGRLTDGQGRITVDFRNSVLIIMTS 718
Qy 720 NLGAHELLAGLTGKVTMEYVARDCVMEVRKHPPELNLRLDELIVFDPDLSHDQLRKVARL 779
Db 719 NLGAQLQNMDTSPKAYEVAQTVQVMEVRKHPPELNLRLDELIVFDPDLSHDQLRKVARL 778
Qy 780 QMKDVAVRALAEGRVALAVTDAALDYTLAESYDPVYCARPLRRMWEKKVVTLSKMTVREE 839
Db 779 ITEELYGLKQKQIRVSLREKXVYVLESFADMGARFRRWVEKNITTELSRMIISOE 838
Qy 840 IDENSTVYI 848
Db 839 LSPNSTVKV 847
RESULT 13
AAM51668
ID AAM51668 standard; Protein; 867 AA.
XX AC AAM51668;
XX 14-JAN-2002 (first entry)
XX Leishmania major 100 kDa heat shock protein (GenBank: CAA86116).
DE Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
KW Arabidopsis thaliana.
XX Leishmania major.
XX WO200170923-A2.
XX 27-SEP-2001.
XX 20-MAR-2001; 2001WO-US08836.
PR 20-MAR-2000; 2000US-190769P.
PR 18-APR-2000; 2000US-198116P.
XX (ARCH-) ARCH DEV CORP.
PI Lindquist S, Queitsch C, Vierling E;
DR WPI; 2001-639123/73.
DR N-PSDB; AAI66084.
XX Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -
Claim 2; Page -; 91pp; English.
XX The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAM51651-AAM51671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or plant, an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.
CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.
CC

Sequence 867 AA;

Query Match 50.3%; Score 2289; DB 22; Length 867;
 Best Local Similarity 54.3%; Pred. No. 1.3e-161;
 Matches 461; Conservative 151; Mismatches 229; Indels 10; Gaps 5;
 3 PEKFTHTKNTTATATAHELAVNAGHAQFTPLHLAGALISDPTGIPPOAISSAGGENAQAOS 62
 6 PE-WKQAASDLMAAALARKKANGVLDPLHAYVMFEDENSASRAVRKLG---AASV 60
 63 ERVINQALKKPSQPPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQIMGLS 122
 61 KQGLARVAIDAIQNPAPTPRPNSDMRWMTAEQERAAALGDTLMAADHFLALHESKE 120
 123 IRDLNEVGAVATARKVSEVEKLRGKGVKVESASGDTNFQALKTYGRDLVEQA--GKLD 180
 121 VGRILDAAGAKGKAIRATILEWR--KGKKTISDFQDNDYESLNKYAVDLCKQAEKGKLD 178
 181 VIGRDEIRRVRLSRRTKNNPVLIGEPGVGKTAVVEGLAQRIIVKGDVPSNLTDLVRLIS 240
 179 VIGRADEILRTIVLSRRTKNNPVLIGEPGVGKTAVVEGLAQRIIVKGDVPSNLTDLVRLIS 238
 241 LDMGALVAGAKYRGEFEERLKVILVEDEAGKVLFTDEIHLVIGAGTEGMDAANLF 300
 239 LDMGALVAGAKYRGEFEERLKVILVEDEAGKVLFTDEIHLVIGAGTEGMDAANLF 298
 301 KPMALARGQLRCIGATTLEBYRKVKVEKDAFERFQVVAEPSPDTTISILRGLKEKVEG 360
 299 KPLLAGDVRTIGATTLEBYRKVKVEKDAFERFQVVAEPSPDTTISILRGLKEKVEG 358
 361 HGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVOLDSPBEIDNLERK 420
 359 HGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVOLDSPBEIDNLERK 418
 421 MOLETELHAREKQKAKARLIEVRKELDLRDLKPLTKYRKEKERIDEIRLKKR 480
 419 ROLETELHAREKQKAKARLIEVRKELDLRDLKPLTKYRKEKERIDEIRLKKR 478
 481 BELMFSLOBAERRYDLARAADLRYGAIQEVESAIQAL-EGTSSBENVMLTENVGPHLAE 539
 479 DEKKVLEAERMRDMETAADLKRVIPILQDRIRSLKAEIKQKATMLQGTVTGTDIAT 538
 540 VVSRWTGIPVTLQNEKERLIGLADRLHKKRVGQONAVNAVSEAILRSRAGLGRAOQPT 599
 539 VVSRWTNIPVTKLSQTERERLHLADQLHLRVKGDQEAASRAILRSRAGLGRASDEPT 598
 600 GSFLFGPTGVGKTELAKALABOLFDENLVRIDMSYMEQSHSVRLIGAPPGVVGHEE 659
 599 GSFLFGPTGVGKTELAKALABOLFDENLVRIDMSYMEQSHSVRLIGAPPGVVGHEE 658
 660 GQQLTEAVRRRPPYCVILFDEVEKAHVAVENTILQVLDGRLTDGQRTVDPRNSVIIMTS 719
 659 GQQLTEAVRRRPPYCVILFDEVEKAHVAVENTILQVLDGRLTDGQRTVDPRNSVIIMTS 718
 720 NLGAREHLLAGTGTMTVEARDVCVMREKVRKHPRLNRLDEIVFDPDLSDHQLRKVARL 779
 719 NLGAYQLQNMDDTSPKPYEVAQVQVMEVRFKFRPEFINRLDDIILFRSLGKEMTGDI 778
 780 QMKDVAERLARGVALVATDAALDYLAESVDPVVGARPIRRMWEKVVITSLSKVREE 839
 779 ITEELNGLRKGQSIRVSLTEBAKQVLESADFADMGARPLRRRWEKVVITSLSKVREE 838
 840 IDENSTVYI 848
 839 LSPNSTVKV 847

ULT 14
 77915
 ABG77915 standard; Protein; 872 AA.
 ABG77915;
 77915

05-NOV-2002 (first entry)
 High level promoter polypeptide #19.
 High level promoter; Cyanobacterium; UV-B light; green alga; plant;
 sunlight biomaterial.
 Synechocystis sp. PCC6803.
 WO200261098-A2.
 08-AUG-2002.
 30-JAN-2002; 2002WO-US03926.
 30-JAN-2001; 2001US-264925P.
 (DUPO) DU PONT DE NEMOURS & CO E I.
 Huang LL, Larossa RA, McCluskey MP;
 WPI; 2002-619256/66.
 N-ESDB; ABS63272.
 Regulating gene expression of a coding region in Cyanobacterium, useful
 in identifying highly expressed or UV responsive genes and their
 promoters, comprises culturing the transformed cell in log phase or in
 the presence of UV-B light -
 Claim 4; Page 83-85; 86pp; English.
 The invention relates to a method for regulating gene expression of a
 coding region in a Cyanobacterium comprising culturing the transformed
 Cyanobacterium in the log phase or in the presence of UV-B light, where
 the promoter region is activated and the coding region is expressed. The
 method is useful for identification of highly expressed genes or UV
 responsive genes, and their corresponding promoters. These genes and
 promoters are useful for constructing expression vectors in
 Cyanobacteria, green algae or plants and for the production of
 biomaterials from sunlight. This sequence represents a high level
 promoter polypeptide of the invention.
 Sequence 872 AA;
 Query Match 50.3%; Score 2288; DB 23; Length 872;
 Best Local Similarity 52.6%; Pred. No. 1.3e-161;
 Matches 460; Conservative 167; Mismatches 229; Indels 18; Gaps 7;
 2 NPEKFTHTKNTTATATAHELAVNAGHAQFTPLHLAGALISDPTGIPPOAISSAGGENAQAOS 61
 5 DPNKFTKAWALAKTPEIAKQHRQOQIETHELLSALL-EQGLATSIFNKAGA-----S 58
 62 AERV---INQALKPSQPPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQIMGLL 118
 59 IPRVNDQVNSFIAQPKLSNPSESIYLGRLSKLDNAEIAKSKYGDVYISTEHLMAAYG 118
 119 EDSQI-RDILNEVGAVATARKVSEVEKLRGKGVKVESASGDTNFQALKTYGRDLVEQA-- 175
 119 QDRILGNLYREIGLTENKLAIIKQIRGTQ--KYTDQNPCKEYSELEKYGRLDTLARE 176
 176 GKLDPVIGRDEEIRRVRLSRRTKNNPVLIGEPGVGKTAVVEGLAQRIIVKGDVPSNLT 235
 177 GKLDPVIGRDEEIRRVRLSRRTKNNPVLIGEPGVGKTAVVEGLAQRIIVKGDVPSNLT 236
 236 VRLISLDMGALVAGAKYRGEFEERLKVILVEDEAGKVLFTDEIHLVIGAGTEGMD 295
 237 RKLLISLDMGALVAGAKYRGEFEERLKVILVEDEAGKVLFTDEIHLVIGAGTEGMD 296
 296 AANLKPMLARGQLRCIGATTLEBYRKVKVEKDAFERFQVVAEPSPDTTISILRGLK 355
 297 AGNLLKPMALARGQLRCIGATTLEBYRKVKVEKDAFERFQVVAEPSPDTTISILRGLK 356
 356 EKYEGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVOLDSPBEIDN 415

```

357 ERYEVHGVKADSAVAAAMLSNRYISDRFLPKAIDLVDVDAKAKMEITTSKPELDE 416
416 LERKEMOLEIEHLEHAKERKOKASKARLIEVKEKELDLKLOPLTKMYRKEKERIDIRR 475
417 VDRKILQEMERLSLQRENDSASERLEKLEKELADFKKEOSKLGWQSEKTVIDQIRT 476
476 LKQREELMFLSLOAERRYDLARAADLYGAIQVEGSAIAQEGTSSEENV---MLTEN 531
477 VKETIDVNLBIQQAQRDYDKAAELQYKGLTDLQVQVEALEFQLAEQOTSGKSLREE 536
532 VGPHEIAEVRSGWTGPIVTRIGQNEKERLIGLARLHKRVVQGNQAVNAVEATILSRAG 591
537 VLESIDIAEIIKWTGIPITKLVSEKSKLLHLEDELHRSRVIGQDEAVTAVAAEIORSRAG 596
592 LGRAQOPTGSLFPGTGVGKTELAKALAEOLFDDENLLVRIDMSEYMEQHSVRLIGAP 651
597 LSDNRPFTASIFLPGTGVGKTELAKALAKNLFTDEALVRIDMSEYMEKHAVSRIMGAP 656
652 PGYVHBEGGQLTEAVRRPVCVILLFDEVEKARHVAFTNLLQVLDGRLTDGQRTVDFR 711
657 PGYVYBEGGQLTEAIRRPVSIVLFDIEKAGGVFNWMLQILDGRLTDGQRTVDFR 716
712 NSVIMTSLNLAELHLAGLTKVMTMEVARDVCMREVRKHPRELLNRLDRIVVPDPLSHD 771
717 NTIIMTSLNLSQVILDVAGDSDRYEEMSRVMDVMRENERPEFLNRVDETIIFHGLQKS 776
772 QLRKVARLQMKDVAVALAERGAVALAVTDALDYLAESYDPVYVGARPIREMEKVVTEL 831
777 ELRSIVQIQSLATREBEQKLTUKTKALDPLAAVGYDPVYVGARPLKRAVOKYLETAI 836
832 SKMVREIDENSTVYIDAGAGDLVYRVESGGLV 865
837 AKGILRGDKPGETIVVDDETDERLSFTSLRGDLV 870

```

JUL 15
199337

ABB49337 standard; Protein; 866 AA.

ABB49337;

05-FEB-2002 (first entry)

Listeria monocytogenes protein #2041.

Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
vitamin B12; bacterial infection; disease.

Listeria monocytogenes.

WO200177335-A2.

18-OCT-2001.

11-APR-2001; 2001WO-FR01118.

11-APR-2000; 2000FR-0004629.

(INST) INST PASTEUR.

Ruchrieser C, Prangeul L, Couve E, Rusniok C, Faihi H, Dehoux P;
Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
Rose M, Voss H;

WPI; 2002-010914/01.

Genomic sequence for Listeria monocytogenes, useful e.g. for treatment

and prevention of Listeria and related bacterial infections, and
related polypeptides

Claim 6; SEQ ID No 2042; 192pp; French.

The present invention relates to the genome sequence of Listeria
monocytogenes BGD-e (see ABA03041). The genome sequence and fragments of
it are useful for selecting probes and primers for detecting genes in L.
monocytogenes and related organisms, and for studying genetic
polymorphisms and other genomes. The present sequence is a protein
encoded by the genome sequence of the present invention. Proteins
expressed from the genome sequence are useful for raising specific
antibodies. Identification of L. monocytogenes and related organisms, and
for biosynthesis and biodegradation, especially biosynthesis of vitamin
B12. The genome sequence and proteins encoded by it are also useful for
selecting compounds that regulate gene expression and cell replication
and modulate L. monocytogenes-related diseases. In addition, the genome
sequence and proteins encoded by it are useful in pharmaceutical and
vaccine compositions for the treatment or prevention of infections by L.
monocytogenes and related organisms.

Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIFO
at ftp.wipo.int/pub/published_pct_sequences.

Sequence 866 AA;

Query Match 49.2%; Score 2236.5; DB 23; Length 866;
Best Local Similarity 51.5%; Pred. No. 8.9e-158;
Matches 448; Conservative 181; Mismatches 224; Indels 17; Gaps 9;

```

QY 1 MNPEKFTKNTETIATAHELVNAGHAQFTPLHLAALISDPTGIFPQAISAGGENAAQ 60
DB 1 MDLQFTQVQQTIAQAQNLAIASEHQIDVAHVFKVLLTRES--FAKRVVDV-AEVDTD 57
QY 61 SAERVINGALKLP--SQSPPPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQLIMGL 118
DB 58 ALQKVIENTLEKIPVVGSGVNYGQMSQALFQLMRDAKEQQLDQDQFVSTEHLILAVM 117
QY 119 E--DSQIRDLNEVGAVATARKVSEVKLGKGGKYSASGDTNFOALKTYGRDLVEQ-- 174
DB 118 DQKSNPITAEKNOHKAQKQKAEAILKIRG--GKRVTSQNAENYEALTKYGRDLVAER 175
QY 175 AGKLDPVIGRDEEIRRVVRLISRTKNNPVLIGPGVGTAVVEGLAQRIYKGVDPNSLT 234
DB 176 SKGLDPVIGRAEINRVIRLSRTKNNPVLIGPGVGTAVVEGLAQRIYKGVDPNSLT 235
QY 235 DVRLISLDMKALVAGAKYRGEPFERLSKVLKEVEDABGKVLFDIETHLVLGAKTEGSM 294
DB 236 DKTIIISLDIGSLIAGAKYRGEPFERLSKVLKEVEDABGKVLFDIETHLVLGAKTEGSM 295
QY 295 DAANLFKPMALARGQLRCIGATTLSEYRYKVEKDAAFRRFPQVYVAPSVDDTISIRGL 354
DB 296 DAGNMLKPMALARGELHCIGATTLDEYRYKVEKDAAFRRFPQVYVAPSVDDTISIRGL 355
QY 355 KEKYGHHGVRIQDRALINAAQLSARYITGRHLDPKADLDVDEACANRVVOLDSPQSEID 414
DB 356 KERFEIHHGVNIHDNALVAASLSNRVITDRLPDKADLDVDEACATIRVEIDMPSELD 415
QY 415 NLERKRMOLIEIHLALEREKDKASKARLIEVRKELDLRLDKLQPLTKYRKEKERIDEIR 474
DB 416 EVTRKVMQLETEEAALKKEKDPASERRELETLQRELAQVKEANQMSKSEKNEISKIR 475
QY 475 RLKQKRELMESLOEASRRYDLAADAALRGATQVESATAQLEGTSSE----ENVMLTE 530
DB 476 EVRFQIDHRLHELESEANNVNLKAAELRHGRIPAVEKELLELEAEENREKTAQEDRLQ 535
QY 531 NVGPEHIAEVVSRWTGIPVTRIGQNEKERLIGLARLHKRVVQGNQAVNAVEATILSRSA 590
DB 536 EVTENEIAEIVGRWTGIPVTKLVEGEREKLLKADVLHQVIGQDDAVQLVDAVLARA 595
QY 591 GLGRAQOQPTGSLFLGPTGVGKTELAKALAEOLFDDENLLVRIDMSEYMEQHSVRLIGA 650
DB 596 GIKDPKPIGSGFIFLPGTGVGKTELAKALAFNMFDSSEHDMIRIDMSEYMEKHSVRLVGA 655

```


651 PPGYVGEHGGQTEAVRRPFCVILFDEVEKAVAVFTLLQVLDGRLTDCQGRVDF 710
 656 PPGYVGEHGGQTEAVRRPFCVILFDEVEKAVAVFTLLQVLDGRLTDCQGRVDF 715
 711 RNSVLIIMTSLGAEHLLAGL-TGKVTMEVARDCCVREVRKHFPPELNLRLDEIVVDPPLS 769
 716 KNTVLIIMTSLGAEHLLAGL-TGKVTMEVARDCCVREVRKHFPPELNLRLDEIVVDPPLS 774
 770 HDQLRKVARLQMKDVAVRLAERGVALAVTDALDYILABSYDPVYGARPIRMMKKVVT 829
 775 LADIKGIVEKVLVEQLRLADQBITTISDDAKAFIABEAYDPVYGARPIRMMKKVVT 834
 830 ELKMWVREIDENSTVYIDAGADLVYRV 859
 835 PLAREIVSGKIMPHSSVEIDLADKEFTKV 864

arch completed: February 13, 2004, 01:30:55
 o time : 77 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 13, 2004, 01:10:25 ; Search time 47 seconds
(without alignments)
1864.036 Million cell updates/sec

le: US-09-812-350-17

fect score: 4550

quence: 1 MNPEKFTHTNETHATAHEL.....VKOMRIEIEDDDNEEMIED 911

ring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

ched: 283308 seqs, 96168682 residues

al number of hits satisfying chosen parameters: 283308

imum DB seq length: 0

imum DB seq length: 2000000000

t-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ult No.	Score	Query Match	Length	DB ID	Description
1	4545	99.9	911	2 F96771	heat shock protein
2	4003	88.0	911	1 T07807	endopeptidase Clp
3	2807	61.7	831	2 D71409	probable endopepti
4	2431	53.4	668	2 G85160	heat shock protein
5	2294	50.4	931	2 AP3276	ATP-dependent clp
6	2288	50.3	872	2 S76197	endopeptidase Clp
7	2258.5	49.6	865	2 D97018	ATPase with chaper
8	2242.5	49.3	866	2 A11720	endopeptidase Clp
9	2236.5	49.2	866	2 AF1350	endopeptidase Clp
10	2234.5	49.1	905	2 T39572	probable proteinas
11	2224	48.9	859	2 C87358	hypothetical prote
12	2218.5	48.8	835	2 AD2441	endopeptidase Clp
13	2212.5	48.6	862	1 E35905	endopeptidase Clp
14	2204	48.4	857	2 A10831	ClpB protein (heat
15	2199	48.3	874	2 AC3070	ATP-dependent Clp
16	2199	48.3	887	2 F98216	endopeptidase Clp
17	2193	48.2	857	1 D35905	endopeptidase Clp
18	2193	48.2	857	2 G91060	heat shock protein
19	2193	48.2	861	2 E85905	heat shock protein
20	2182.5	48.0	898	2 S76431	endopeptidase Clp
21	2168.5	47.7	858	2 D71711	endopeptidase Clp
22	2166	47.6	968	2 T51523	clpB heat shock pr
23	2163.5	47.5	857	2 C97707	clpB protein (impo
24	2162.5	47.5	856	1 F64098	endopeptidase Clp
25	2160	47.5	857	2 A10397	Clp ATPase [import
26	2158.5	47.4	848	2 C70834	probable endopepti
27	2158.5	47.4	861	2 D82814	ATP-dependent Clp
28	2153.5	47.3	853	2 T36551	probable ATP-depen
29	2150.5	47.3	848	2 G87220	heat shock protein

30	2141.5	47.1	869	2 H89864	hypothetical prote
31	2133.5	46.9	880	2 AC2096	endopeptidase Clp
32	2133	46.9	854	2 D83077	ClpB protein PA454
33	2119.5	46.6	875	2 G75442	ATP-dependent Clp
34	2104.5	46.3	859	2 F81078	clpB protein NMB14
35	2104	46.2	878	2 G71371	probable endopepti
36	2103.5	46.2	859	2 F81863	ClpB protein NMA16
37	2099.5	46.1	857	2 E86815	clpB protein (impo
38	2086	45.8	857	2 A82290	clpB protein VC071
39	2069	45.5	874	2 G84644	probable ATP-depen
40	2039	44.8	860	1 C35905	endopeptidase Clp
41	1943	42.7	867	2 A81707	ATP-dependent Clp
42	1938	42.6	867	2 D71555	endopeptidase Clp
43	1927.5	42.4	857	2 F81396	ATP-dependent Clp
44	1912	42.0	866	2 E72113	clp proteinase ATP
45	1912	42.0	866	2 H86508	Clp proteinase ATP

ALIGNMENTS

RESULT 1

F96771

heat shock protein 101, 13093-16240 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C;Accession: F96771

R;Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Chin, C.W.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maili, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: F96771

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-911 <STO>

A;Cross-references: GB:AE005173; NID:g6939223; PIDN:AAE31725.1; GSPDB:GN00141

C;Genetics:

A;Gene: F1017.2

A;Map position: 1

C;Superfamily: endopeptidase Clp ATP-binding chain

Query Match 99.9%; Score 4545; DB 2; Length 911;
Best Local Similarity 99.9%; Pred. No. 6.6e-194;
Matches 910; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MNPEKFTHTNETHATAHELAVNAGHAQFTPLHLAGALISDPTGTFPQAISSAGGENAAQ	60
DB	1	MNPEKFTHTNETHATAHELAVNAGHAQFTPLHLAGALISDPTGTFPQAISSAGGENAAQ	60
QY	61	SAERVINGALKKLPQSPPPDDIPASSSLIVIRRAQAQKSRGDTHLAVDQIMGLLED	120
DB	61	SAERVINGALKKLPQSPPPDDIPASSSLIVIRRAQAQKSRGDTHLAVDQIMGLLED	120
QY	121	SGIRDLLNEGVATARVKSVEKLRGEGKKEKVESAGDTNFAQKTYGRDLVEQAGKLDLP	180
DB	121	SGIRDLLNEGVATARVKSVEKLRGEGKKEKVESAGDTNFAQKTYGRDLVEQAGKLDLP	180
QY	181	VIGRDEEIRRVVRLISRRTKNPVLIIGEPGVGKTAVVEGLAQRIKVGSDVNSLPDVLRLIS	240
DB	181	VIGRDEEIRRVVRLISRRTKNPVLIIGEPGVGKTAVVEGLAQRIKVGSDVNSLPDVLRLIS	240
QY	241	LDMGALVAGAKYRGFEERLKSVLKEVEDARGKVLFDTHLVGLGAGKTEGSDAANLF	300
DB	241	LDMGALVAGAKYRGFEERLKSVLKEVEDARGKVLFDTHLVGLGAGKTEGSDAANLF	300
QY	301	KPMLARGQLRCIGATTLEERYKVEKDAAFERFRQQVYVAPSPVPTISILRLGKKEYG	360

RESULT 3
D71409
probable
N; Altern
N; Contain
C; Specie

ariety: columbia
ate: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 19-Jan-2001
cession: D71409
edler, H.; Wedler, E.; Kambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Giel
agh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
are 391, 485-488, 1998
hthors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
ft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; And
Chalwatzis, N.
le. Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
ference number: A71400; MUID:98121113; PMID:9461215
cession: D71409
atus: preliminary; nucleic acid sequence not shown; translation not shown
molecule type: DNA
esidues: 1-831 <BBV>
ss-references: GB:297336; NID:G2244788; PIDN:CAB10246.1; PID:G2244823
netics:
p position: 4COP9-4G3845
nction:
scription: allows clip to hydrolyze polypeptides and proteins, probably by a chaperon
tivity; ATP hydrolysis is required for Clip hydrolysis of proteins but not of smaller
family: endopeptidase Clip ATP-binding chain
ords: ATP; hydrolase; molecular chaperone; nucleotide binding; P-loop
2-179/Region: nucleotide-binding motif A (P-loop)
0-245/Region: nucleotide-binding motif B
1-578/Region: nucleotide-binding motif A (P-loop)
0-645/Region: nucleotide-binding motif B
8/Binding site: ATP (Lys) #status predicted
7/Binding site: ATP (Lys) #status predicted
ery Match 61.7%; Score 2807; DB 2; Length 831;
st Local Similarity 64.1%; Pred. No. 5e-117;
tches 585; Conservative 108; Mismatches 129; Indels 90; Gaps 11;
1 MNPEKFTKNTETATAHAVNAGHAQFTPLHLAALISDPTGIPFOAISSAG-GENAA 59
1 MNDLKFDPNVKLIILASARSAMSLSHGQVTPHLGVTLISDLTSVFYRAITSAGDGDISA 60
60 QSAERVINQALKKLPSSQPPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQIMGLLE 119
61 QSVNVINQSLYKL-----TKRNLGDTKGVAVLVISLLE 95
120 DSQIRDLLENGVATARKVSEKLRGKGGKVESAGDPTNFQALKTGYGRDLVEQAGKLD 179
96 DSQISDLVKEAGVVPKVESEKLR-----GEVILRALTKYGTDLVEQAGKLD 144
180 PVIGRDEEIRRVVIRLSRRTKNNPVLIGBPGVKTAVVEGLAQRIKGVDPNSLTDVRLI 239
145 PVIGRHEIRRVIEVLSRRTKNNPVLIGBPGVKTAVVEGLAQRIKGVDPINLTGVKLI 204
240 SLDMGALVAGAKYGRFEERLKSVLKEVEDAEQKVLIFIDEIHLVLAGKTEGSDMAANL 299
205 SLEFGAMVAGTTLRGQFEERLKSVLKAVEAQKVLVFIDEIHLGACAKSAGSTDAKL 264
300 FKPMALRGQLRCIGATTLEBYRYKVEKDAAFERRFOQVYVABSPVDDTISILRLKEKYE 359
265 LKPMALRGQLRFIGATTLEBYRTHVEKDAAFERRFOQVYVABSPVDDTISILRLKEKYE 324
360 GHGVRIOQRALINAAQLSARYITGRLHPDKAIDLVDACANVRVQLDSQPEINLERK 419
325 GHGVRIOQRALVLSAQLSERYITGRRLPKAIDLVDSECAHVKAQLDQPEIDSLERK 384
420 RMQLEIHLALEREK-DKASKARLIEVRKELDLRLKQPLTKYKKEKERIDEITRLKQ 478
385 VMQLEIEIHALEKEDKASEARLSEVRKELDLRLKLEPLIKYKKEKKIINETRLKQ 444
479 KREELMFSQAEARRVLDARAADLYGATQEVESAIAQLEGTSSEENVMLTENVGPEHIA 538
445 NRDDLMLAQAEARQHDVPAKAVLYKYGAQEVESAIAKLE-KSADKNVMLTETVGPENIA 503
539 EVVSRWTGIPVTRLGONEKERLIGLADRLHKRVGQNAVSEAILRSRAGLGRAQOP 598

Db 504 EVVSRWTGIPVTRLGONEKRLISLADKLHERVVQDEAVKAVAAAILRSRVLGRPOOP 563
QY 599 TGSFPLFGTGVGKTELAKALAEQIFDDENLLVRIDMGEYMEQHSVRLIGAPPG-YVGH 657
Db 564 SGSFPLFGTGVGKTELAKALAEQIFDSENLLVRIDMGEYNDKFSVNLIGAPPGYIGH 623
QY 658 EGGGOLTRAVRRPVCVILFDRVEKARHVAVENTLLQVLDGDLTDCQGTVDFRNSVLIIM 717
Db 624 EGGGOLTRAVRRPVCVILFDRVEKARHVAVENTLLQVLDGDLTDCQGTVDFRNSVLIIM 673
QY 718 TSNLGAHLLAGLTKGVIMEVARDCVMREVRKHFPELINRLIDEIVDFPLSHDQLRKVA 777
Db 674 TSNLGAHLLVSLTCEMTQVARDNAMDKAKHFEPLNRLIDEIVDFPLSHDLAKIV 733
QY 778 RLOMKOVAVRLAERGVALLAVTDAALDYILAESYDPVYGARPIRMWMEKKVVTLSKVVVR 837
Db 734 QLOQKXNV-----PIRWLERKVVTDISMIVR 761
QY 838 REIDENSTVYIDAGAG--DLVVRVESGGLVDASTGKSDVLIHIANGPKRSDAAQAV--- 892
Db 762 BEIGDSDSIVCIDVNEAKTDLVYRDKNVFVKIE--QTLDVVVIHSGNKGKGRNDEHDHIVTL 819
QY 893 -KKMRIRIEDD 903
Db 820 TKKIKNEVVVID 831
RESULT 4
G85160
heat shock protein like [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
C:Accession: G85160
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: G85160
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-868 <STO>
A:Cross-references: GB:NC_001268; NID:G5302773; PIDN:CAB46061.1; GSPDB:GN00140
C:Genetics:
A:Gene: dl3375w
A:Map position: 4
C:Superfamily: endopeptidase Clip ATP-binding chain
Query Match 53.4%; Score 2431; DB 2; Length 668;
Best Local Similarity 70.8%; Pred. No. 1.7e-100;
Matches 499; Conservative 76; Mismatches 90; Indels 40; Gaps 6;
QY 1 MNPEKFTKNTETATAHAVNAGHAQFTPLHLAALISDPTGIPFOAISSAG-GENAA 59
Db 1 MNDLKFDPNVKLIILASARSAMSLSHGQVTPHLGVTLISDLTSVFYRAITSAGDGDISA 60
QY 60 QSAERVINQALKKLPSSQPPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQIMGLLE 119
Db 61 QSVNVINQSLYKL-----TKRNLGDTKGVAVLVISLLE 95
QY 120 DSQIRDLLENGVATARKVSEKLRGKGGKVESAGDPTNFQALKTGYGRDLVEQAGKLD 179
Db 96 DSQISDLVKEAGVVPKVESEKLR-----GEVILRALTKYGTDLVEQAGKLD 144
QY 180 PVIGRDEEIRRVVIRLSRRTKNNPVLIGBPGVKTAVVEGLAQRIKGVDPNSLTDVRLI 239
Db 145 PVIGRHEIRRVIEVLSRRTKNNPVLIGBPGVKTAVVEGLAQRIKGVDPINLTGVKLI 204
QY 240 SLDMGALVAGAKYGRFEERLKSVLKEVEDAEQKVLIFIDEIHLVLAGKTEGSDMAANL 299
Db 205 SLEFGAMVAGTTLRGQFEERLKSVLKAVEAQKVLVFIDEIHLGACAKSAGSTDAKL 264
QY 300 FKPMALRGQLRCIGATTLEBYRYKVEKDAAFERRFOQVYVABSPVDDTISILRLKEKYE 359

265 LKPLMARGQLRFGATTTLEBYRTHVEKDAAFERFQOVFAEPSPDTISILRLKEKYE 324
 360 GHGVRVQDRALNAAQLSARYITGRHLDPKADLDVDEACANVRVOLDSPQPEIDNLERK 419
 325 GHGVRVQDRALNAAQLSARYITGRHLDPKADLDVDEACANVRVOLDSPQPEIDNLERK 384
 420 RMQLEIEHALEREK-DKASKARLIEYRKELDLRLKLOPLTKYKKEKERIDEIRRLKQ 478
 385 VMQLEIEHALEREK-DKASKARLIEYRKELDLRLKLOPLTKYKKEKERIDEIRRLKQ 444
 479 KREELMESLOEAERYDRLARAADLRGATQEVESAIQLEGTSSEENVMLTENVGP 538
 445 NRDLMLAQEAERQHDVPKAAVLYKGAIQEVESAIQLEGTSSEENVMLTENVGP 503
 539 EVVSRWTGIPVTRLQNGEKERLIGLADRLHKKRVVQGNQAVNAVSEAILRSRAGLGR 598
 504 EVVSRWTGIPVTRLQNGEKERLIGLADRLHKKRVVQGNQAVNAVSEAILRSRAGLGR 563
 599 TGSFLFLGPTGVGKTELAKALAEQDFDENLVRIDMSYMEQHSVSRLIGAPPG-YVGH 657
 564 SGSFLFLGPTGVGKTELAKALAEQDFDENLVRIDMSYMEQHSVSRLIGAPPGYVGH 623
 658 EGGQTEAVRRRPPYCVILFDEVEKAHVAVFNTLLQVLDGRLTD 702
 624 EGGQTEPVRRRPPYCVILFDEVEKAHVAVFNTLLQVLDGRLTD 668

RESULT 5
 276
 3-dependent clp proteinase, ATP-binding chain clpb BMEI0195 [imported] - Brucella meli
 Species: Brucella melitensis
 Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
 Accession: AF3276
 V.G.; Kapratral, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,
 Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
 c. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
 Reference number: AD3252; PMID:11756688
 Accession: AF3276
 Status: preliminary
 Molecule type: DNA
 Residues: 1-931 <KUR>
 Cross-references: GB:AE08917; PIDN:AAL51377.1; PID:gl7982078; GSPDB:GN00190
 Experimental source: strain 16M
 Gene: BMEI0195
 Map position: 1
 Superfamily: endopeptidase Clp ATP-binding chain

Query Match 50.4%; Score 2294; DB 2; Length 931;
 Best Local Similarity 53.5%; Pred. No. 3e-94;
 Matches 471; Conservative 151; Mismatches 244; Indels 14; Gaps 7;

1 MNPEKFTKNTETATHELVNAGHAQTPPLHLAGALISDPTGFFPQAISSAGGENAAQ 60
 58 MNIEKYTERVRFIOQAFTALSSNGQFTPHILKVLDDDEGLAASLVRAGER--VG 115
 61 SAERVINQALKLPSQPPDPDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLMGLL-- 118
 116 DVMRGLQSALEKLPKVGSGNDQLYLSQPLAKVFSLAELASAKAGDSFTVERLLTALAME 175
 119 EDSQJRDLLNEGVATARKVSEKLRGKGGKVESASGDTNFQALTYGRDLVE--QAG 176
 176 KSAKTEILSAAAGVTPTALNRVINDMR--KGTADSASAEENYDALKKYARDLTEDARAG 233
 177 KLDPVIGRDEERVRVRLSRKKNPVLIGPPGVKTAAGVGAQRVKGDPVNSLTDV 236
 234 KLDPVIGRDEERVRVRLSRKKNPVLIGPPGVKTAAGVGAQRVKGDPVNSLTDV 293
 237 RLISLDMGALVAGAKYRGFEERLKSVLKEVEDARGKVFILFIDEHLVLGAGKTEGSMDA 296
 294 QLMALDMGALVAGAKYRGFEERLKSVLKEVEDARGKVFILFIDEHLVLGAGKTEGSMDA 353

QY 297 ANLKPMALARGQLRCIGATTTLEBYRTHVEKDAAFERFQOVFAEPSPDTISILRLKE 356
 DB 354 SNLLKPALARGELCHVGATTTLEBYRTHVEKDAAFERFQOVFAEPSPDTISILRLKE 413
 QY 357 KYEGHGVRIQDRALNAAQLSARYITGRHLDPKADLDVDEACANVRVOLDSPQPEIDNL 416
 DB 414 KYEGHGVRIQDRALNAAQLSARYITGRHLDPKADLDVDEACANVRVOLDSPQPEIDNL 473
 QY 417 ERKMOLEIEHALEREK-DKASKARLIEYRKELDLRLKLOPLTKYKKEKERIDEIRRL 476
 DB 474 DRIMQLEIEHALEREK-DKASKARLIEYRKELDLRLKLOPLTKYKKEKERIDEIRRL 533
 QY 477 KQKEELMFISQEAERYDRLARAADLRGATQEVESAIQLEGTSSEENVMLTENVGP 534
 DB 534 KQKEELMFISQEAERYDRLARAADLRGATQEVESAIQLEGTSSEENVMLTENVGP 590
 QY 535 BHIAEVSRWTGIPVTRLQNGEKERLIGLADRLHKKRVVQGNQAVNAVSEAILRSRAGLGR 594
 DB 591 DHVAQVSRWTGIPVTRLQNGEKERLIGLADRLHKKRVVQGNQAVNAVSEAILRSRAGLGR 650
 QY 595 AQQTGSFLFLGPTGVGKTELAKALAEQDFDENLVRIDMSYMEQHSVSRLIGAPPGY 654
 DB 651 PNRPIGSGFIFLPGTGVGKTELAKALAEQDFDENLVRIDMSYMEQHSVSRLIGAPPGY 710
 QY 655 VGHEEGQLTEAVRRRPPYCVILFDEVEKAHVAVFNTLLQVLDGRLTDGQCRVDFRNSV 714
 DB 711 VGHEEGQLTEAVRRRPPYCVILFDEVEKAHVAVFNTLLQVLDGRLTDGQCRVDFRNSV 770
 QY 715 IIMTSNIGASHLLAGLTKVTMEVARDQVVRKPRPELNLRLDEIVVDFDPLSHDQLR 774
 DB 771 IIMTSNIGASHLLAGLTKVTMEVARDQVVRKPRPELNLRLDEIVVDFDPLSHDQLR 829
 QY 775 KVARLOMKDVAVRLAEHGVVALVTDALDVLAEVSDPVYVGAERPIRRMEKKVVTLSXK 834
 DB 830 AIVDIQMRQYLLSDRKITLQEDDAREWLANKGDPAYGARPLKRVIOKQVQDPLAER 889
 QY 835 VVREEDENSTVYIDAGDLVYRVESGGLVDASTGKXSD 874
 DB 890 ILLGDILLGSLVKITAGSDRLNFRPISGAPSAAPERDE 929

RESULT 6
 S76197
 endopeptidase Clp ATP-binding chain B1 - Synechocystis sp. (strain PCC 6803)
 N;Alternate names: ATP-dependent Clp proteinase regulatory chain; protein slr1641
 N;Contains: adenosinetriphosphatase (EC 3.6.1.3)
 C;Species: Synechocystis sp.
 A;Variety: PCC 6803
 C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 19-Jan-2001
 C;Accession: S76197
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 S.
 A;Reference number: S74322; PMID:97061201; PMID:8905231
 A;Accession: S76197
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-872 <KAN>
 A;Cross-references: EMBL:D90914; GB:AB001339; NID:gl653477; PIDN:BAAL8456.1; PID:gl65354
 C;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C;Genetics:
 A;Gene: clpB1
 C;Function:
 A;Description: allows clpP to hydrolyze polypeptides and proteins, probably by a chapero-
 e activity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller
 C;Superfamily: endopeptidase Clp ATP-binding chain
 C;Keywords: ATP; hydrolase; molecular chaperone; nucleotide binding; P-loop
 P;208-215/Region: nucleotide-binding motif A (P-loop)
 P;276-280/Region: nucleotide-binding motif B
 P;611-618/Region: nucleotide-binding motif A (P-loop)
 P;679-683/Region: nucleotide-binding motif B

Species: Clostridium acetobutylicum
Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

QY 772 QLRKVARLQMKDVAVRLAERGVALAVTDAALDYILAESYDPVYGARPIRRWMEKKVUTEL 831

ULT 7
018
use w:
pecies
ute: 1

536 EVTENEIAEIVGRWTGIVTTLVKGEREKLLADLVLHKVIGQDDAVQLVSDAVLRARA 595
 591 GLGRGAQOQTGFLGPGTGVGKTELAKALAEOLFDDENLLVRIDMSEYMEQHSVSRLLGA 650
 596 GIKQPKRPIGSGFIFLPGTGVGKTELAKALAYNMFDSHMFIDMSEYMEKHSVSRLVGA 655
 651 PPGVYGHGEGQLTEAVRRRVCYVILFDEVKAAHVAVNTLLQVLDLDDGRLTDGQGRVDF 710
 656 PPGVYGVGEGQLTEAVRNPYSIVLLDEIEKAHPDVFNILLQVLDLDDGRITDSQGRLLDF 715
 711 RNSVLTMSNLGAEHLIAGL--TGKVTMEVARDVCMREVRKHFPRPELLNRLDEIVVDFPLS 769
 716 KNTVLTMSNTGSNLLLRTERGEISPELSD--VMQILQSEFKPEFLNVRVDDIILFKPLT 774
 770 HDQRLKVARLQMKOVAVRLAERGVALAVTDAALDYILAESYDPVYGARPIRMMEKCVYT 829
 775 LADINGIVEKLVEELQIRLADQEQETITISDNAKAFIABEAYDPVYGARPLKRYIVRVHET 834
 830 ELSKMWVREEIDENSTVVIDAGDLVYR 859
 835 PLAREIVSGKIMPHSSVEIDLQKFEFTFK 864

 RESULT 9
 AF1350
 endonopeptidase Clp ATP-binding chain B (ClpB) homolog clpB [imported] - *Listeria monocytogenes*
 C:Species: *Listeria monocytogenes*
 C:Strain: 27-Nov-2001 #sequence revision 27-Nov-2001 #text_change 14-Dec-2001
 C:Accession: AF1350
 C:ReleaseDate: 27-Nov-2001
 C:Author: J.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
 D.; Jones, L.M.; Karst, U.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
 Science 294, 849-852, 2001
 A:Authors: Kretz, J.; Kuhn, M.; Kunst, P.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma; Ma;
 A:Title: Comparative genomics of *Listeria* species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AF1350
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-866 <GLA>
 A:Cross-references: GB:NC_003210; PID:CAD00284.1; PID:g16411676; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 C:Gene: clpB
 C:Superfamily: endopeptidase Clp ATP-binding chain

 Query Match 49.2%; Score 2236.5; DB 2; Length 866;
 Best Local Similarity 51.5%; Pred. No. 9.5e-92;
 Matches 448; Conservative 181; Mismatches 224; Indels 17; Gaps 9;

 1 MNPKEFTKMTETATAEHLAVNAGHAQFTPLHLAGALISDPTGIFPOAISSAGGENAAQ 60
 1 MDLQKFTQVQQTITADQNLAIASEHQEIDVAHFVKLLTESD--FAKRVYDV-AEVDTD 57

 61 SAERVINQALKLP--SQSPDDIPASSLIKVIRAAQAOKSRGDTHLAVDQLINGLL 118
 58 ALOKVIENTLEKIPVSGSGVNYQAMSQLFQMLDRAEKEQQOLEDFFVSTEHLILAV 117

 119 E--DSQRLDLNVEGVATARVKSEVKLRGEGKKVSASGDTNFOALKTYGRDLVEQ-- 174
 118 DOKSNPTAEIKNOHQKAKQIKAEILKIRG--GRVTSQNAEENYEATUKYGRDLVAEVR 175

 175 AGKLDPVIGRDEEIRRVVRIILSRRTKKNPVLIGPGVGKTAAVVEGLAQRIYKGDVPNSLT 234
 176 SGKLDPVIGRDAEIRNVRIILSRRTKKNPVLIGPGVGKTAIVEGLAQRIYKQDVEGLK 235

 235 DVRLISLDGMALVAGKYRGEPERLSVYLKEVEDAECKVILFDIEHLVLGCKTQCSM 294
 236 DKTIIISLDIGSLIAGKYRGFEERLKAIVQEQVQSQILLFIDEIHTIVGAKTKDGM 295

 295 DAANLFKPLMARGQLRCITGATLLEBYRKYVEKDAAFERRFQGVTVVAPSPDITISLRGL 354

296 DAGNMKEMLARGELHCIGATTLEQYVIEKDALEKERRFOKVLVPEPTVEDTWSILRGL 355
 355 KEKVEGHGVIRQDRAALNAAGLSARYITGTHLPDKALDLVDEACANVRVOLDQSOPEID 414
 356 KERFEIHGVNHIDNALVAASLSNRYITDRFLPDKALDLVDEACATIRVIDSNPSELD 415
 415 NLERKMOLEIELHALEREKOKASKARLIEVRKELDDLRDLKQLPTMKYRKEKERIDEIR 474
 416 EVTRKVMQLEIEEAAKKEKOPASERRLEILORELDYKBEANQMKSEKNEISKIR 475
 475 PLKQKREBELMSLOEAERYDLARAADLYRGAIOVESIAQOEGTSSE-----BNVMLTE 530
 476 EVREQIDHRLHELEAEANNYDLNKAALRHGRIPAVEKELIELEAEENREKTAQDRILQE 535
 531 NVGPBPHIAEVVSRWTGIPVTRLGONEKERLIGLADRLHKKRVGQONQAVNVAESRLSRA 590
 536 EVTENEAIEIVGRWTGIPVTKLVEGEREKLKLDVHLHQVIGQDDAVQLVSDAVLRARA 595
 591 GLGQAQPTGSLFLGPTGVGKTELAKALAQFDDENLLVRIDMSYMEQHSVSRILGA 650
 596 GIKDKPRIGSIFLGPVGVKTELAKALAFNMFDSBDHMIRIDMSYMEKHSVSRILGA 655
 651 PPGVYVGEHGGQLTEAVRRRYPYCVILDEVEKAHVAVNTLLQVLDGRLTDGGRVTDF 710
 656 PPGYIGVEEGQLTEAVRRNPYSIVLDEIEKAHPDVFNILLQVLDGRLTDSQGRILDF 715
 711 RNSVIIMTSNLGASHLAGL--TGKVTMEVARDVVRKHFREPELLNRLDEIVDFPLS 769
 716 KNTVIIMTSNLGASHLAGL--TGKVTMEVARDVVRKHFREPELLNRLDEIVDFPLS 774
 770 HDQURKVARLQMDVAVRLAERGVALAVTDAALDYILAESYDPVYGARPIRRMWEKKVVT 829
 775 LADIKGIVEKLEVLQRLADQETITISDDAKAFIAEEAYDPVYGARPLKRYVIRHVT 834
 830 ELSRWVREBEIDENSTYIADAGDLYVRV 859
 835 PLAREIVSGKIMPHSSVEIDLADKEFTFKV 864

able proteinase subunit - fission yeast (Schizosaccharomyces pombe)
 Species: Schizosaccharomyces pombe
 Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Feb-2001
 Accession: T39572
 Cod. V.: Rajandream, M.A.; Bartell, B.G.; Brown, D.; Churcher, C.M.
 Submitted to the EMBL Data Library, March 1999
 Reference number: 221864
 Accession: T39572
 Status: preliminary; translated from GB/EMBL/DBJ
 Molecule type: DNA
 Sequences: 1-905 <WOO>
 Cross-references: EMBL:AL035637; PINN:CA338512.1; GSPDB:GN00067; SPDB:SPBC16D10.08C
 Experimental source: strain 972h-; cosmid c16D10
 Metrics:
 Name: SPDB:SPBC16D10.08C
 Map position: 2
 Superfamily: endopeptidase Clp ATP-binding chain
 Keywords: ATP; molecular chaperone; nucleotide binding; P-loop
 19-216/Region: nucleotide-binding motif A (P-loop)
 77-282/Region: nucleotide-binding motif B
 21-628/Region: nucleotide-binding motif A (P-loop)
 99-694/Region: nucleotide-binding motif B
 15/Binding site: ATP (Lys) #status predicted
 27/Binding site: ATP (Lys) #status predicted

ery March 49.1%; Score 2234.5; DB 2; Length 905;
 atest Local Similarity 49.5%; Pred. No. 1.2e-91;
 Matches 460; Conservative 158; Mismatches 258; Indels 53; Gaps 11;
 6 FTHKNTNTIATAHELAVNAGHQAQPTPLHLAGALISDP----TGIFPQAATSSAGGEVAAQS 61

Db 6 FTDKAATLSDAYSIAQSYGHSQLTPIHIAAALLSDSDSGTTLRLTIVDKAGD--GQK 63
 QY 62 ARVINQALKKLPSPQPPDDIPASSLLIKVIRRAQAOKSRGDTHLAVDQLIMGLLEDS 121
 Db 64 FERSVTSRVLVPAQPPPEQVTLSPESAKLRLNAHELQKTKQKSYIAQDHFIAVFTKDD 123
 QY 122 QIRDLNEVGVATARVKSVEKLGKKGKVESASGDTNFQALKTYGRDLVBEQA--GKLD 179
 Db 124 TLKSLLAEGAVTPKAFEPFVNVG--NKRIDSNAEBCGFDALNKFTVDLTTELARGQLD 181
 QY 180 PVIGRDEETRRVRLSRRTKKNPVILIGEPGVGKTAIVVEGLAQRIKVGDPVNSLTDVRLI 239
 Db 182 PVIGRDEETRRVRLSRRTKKNPVILIGEPGVGKTAIVVEGLARRIIDDDVPANLSCKLL 241
 QY 240 SLDMGALVAGKVRGEFERLKSVLKEVEDAEAGKVLIFIDEHLVGLGAK--TEGSMDAAN 298
 Db 242 SLDVGLVAGSFRGEFERIKSVLKEVESESTPILFVDEMHLMGAGSGGEGMDAAN 301
 QY 299 LFKPMLARGQLRCIGATTLEERYKRVKEDAAFERFQOVVYAEPSVPDTISILRGLKEKY 358
 Db 302 LKPMARGKLHCIGATTLEAYKYKIEKDAFERAFQILVKEPSIEDTISILRGLKEKY 361
 QY 359 EGHGVRIODRALINAAQLSARYITGRHLPDKAIDLVEACANVRVOLDQSOPEIDNLER 418
 Db 362 EVHGVITISDRALVTAHLASRYLTSRLPDSALDLVDEAAAARVTVRESOPEVLDNLER 421
 QY 419 KRMOLIEHLALEREKOKASKARLIEVRKELDDLRDLKQLPTMKYRKEKERIDEIRRLKQ 478
 Db 422 KLRQURVEITRALERBDEASKERLKAARKEAQVEETRPTRKEYLEKRSGELODAKR 481
 QY 479 KREELMFSLOEAERYDLARAADLYRGAIOVESIAQOEGTSSENV-----526
 Db 482 RLDELKAKAEADERRNDFTLAADLYRGPDLQKRIEYLEQOKKADAEATANAQSGEP 541
 QY 527 MLENVGPBPHIAEVVSRWTGIPVTRLGONEKERLIGLADRLHKKRVGQONQAVNVAESIL 586
 Db 542 LLIDVVGPDQINEIVARWTGIPVTRLKTTEKRLNMEKVLKQVIGQNEAVTAVANAIR 601
 QY 587 RSRAGLGRAQQTGSLFLGPTGVGKTELAKALAQFDDENLLVRIDMSYMEQHSVSR 646
 Db 602 LSRAGLSDPNQPIASFLGPGSGTGKTLTKALASFDDENAMIRIDMSYMEKHSVSR 661
 QY 647 LIGAPPVYVGEHGGQLTEAVRRRYPYCVILDEVEKAHVAVNTLLQVLDGRLTDGGR 706
 Db 662 LIGAPPVYVGEHGGQLTEQLRRRYPYCVILDEIEKAAPVILVTLQVLDGRLTDSQGR 721
 QY 707 TVDFRNSVIIMTSNLGASHILA--GLTGKVTMEVARDVVRKHFREPELLNRLDEIV 763
 Db 722 VYDAKNAVITMISNLGASHILA--GLTGKVTMEVARDVVRKHFREPELLNRLDEIV 780
 QY 764 VFDPLSHDQLRKVARLQMDVAVRLA--ERGVALAVTDAALDYILAESYDPVYGARPIRR 821
 Db 781 IFNRLRRVDIRNIVENRILEVQKRLQSNHRSIKTEVSEADKDLGSGAGSPAYGAPLNR 840
 QY 822 WMEKKVTVTELSRWVREBEIDENSTYIADAGDLYVRVSGGLVDASTGKKSVDLIHTAN 881
 Db 841 VTQNGVLNPMVAILNGQLRKETAHVAVVQNGKIFVKPN-----HEAN 883
 QY 882 GPKRSDAAQVKKMREIEIDDDDEEMIE 910
 Db 884 ANGSAD-----IDMDGIDDDVDNDELE 905

RESULT 11
 C87358
 hypothetical protein C0878 [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: C87358
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

Title: Complete Genome Sequence of *Caulobacter crescentus*.

Reference number: AB7249; MUID:21173698; PMID:11259647

Accession: C87358

Status: preliminary

Molecule type: DNA

Residues: 1-859 <STO>

Cross-references: GB:AE005673; NID:g13422139; PIDN:AAK22863.1; GSPDB:GN00148

Genetics:

Gene: CC0878

Superfamily: endopeptidase Clp ATP-binding chain

Query Match 48.9%; Score 2224; DB 2; Length 859;

Best Local Similarity 51.9%; Pred. No. 3.4e-91;

Matches 446; Conservative 158; Mismatches 246; Indels 10; Gaps 6;

1 MNPEKFTHTKNTETIATAHVLAVNAGHAQPTPLHLAGALISDPTGTFPPQAISSAGGENAAQ 60

1 MNIDLYSDRAKAVGSAQSALARGHQFAPFHILKVLBEKXGSLRALISAGR--PD 58

61 SAERVINOALKKLPSQPPDDIPASSSLIKVIRDAQAQKSRGTHLAVDLQIMGLL-E 119

59 QLDGGVETLLAKTPRVGAGGQLYMKPDTARVFAEAKSAKAAGDAFTVTERLLIAIAKE 118

120 DSQIRDLNVEGVATARVKEVEKLGKGVESASGDTNFQALKTYGRDIVEQA--GK 177

119 GGEAKLFKEAGVSAQSLETAANMR--KGRFADSANAEVGEALRYARDLTAARDGK 176

178 LDPVIGRDEEIRRVRIISRRTPKNNPVLIGEPGVGTAVVEGLAQIRIVKGVDPNSLTDVR 237

177 LDPVIGRDEEIRRTQVLSRRTPKNNPVLIGEPGVGTALVEGLALRVNGDVPESLKDCK 236

238 LLSLDMGALVAGAKYRGFEERLKVLEVEDAEKVLFDIDEHLVLAGKTEGSMDDAA 297

237 LLSLDMGSLIAGAKYRGFEERLKVLEVEDAEKVLFDIDEHLVLAGKTEGSMDDAS 296

298 NLFKPMARGQLRCIGATTLEERYKVKDKDAFFRRFOVVVAEVSPTTILRLGKLEK 357

297 NLLKPMARGQLRCIGATTLEERYKVKDKDAFFRRFOVVVAEVSPTTILRLGKLEK 356

358 YEGHGVRIQDRALINAAQSLARYITGRHLPDKAIDLVEACANRVVOLDQPEIDNLE 417

357 YEVHGVRIQDRALINAAQSLARYITGRHLPDKAIDLVEACANRVVOLDQPEIDNLE 416

418 RKMQLTEHLALEREKQKASVRLIIVRKELDLRLDKLQPLTKMYRKEKRIDEIRLKL 477

417 RLVLQKTEREALSKETDAASKQRLLENLEVEIDDLQFRSDENTARWKAEEKVGAQAAR 476

478 OKRELMFSLQEAERRYDLARAADLRYGAIQVESAIQLEGTSEENVMLTENVGPEHI 537

477 EALDRRLADLANAQRAGDFARAGQIQYGEIPALERRLAEE--AGDTQALTPEVVDABQI 534

538 AEVSRWTGIPVTRIGONEKERLIGLADRLKRVVVGQNAVNAVSEALIRSRAGLGRAQ 597

535 AAVSRWTGIPVPEKMLEGEREKLLKWEDELGRVVGQDEALEAVSDAVRRARAGLDPSK 594

598 PGSGFLFGPTGVGTAKALAEQFDENLLVRIIDMSYMEQHSRLIGAPGVGH 657

595 PGSGFLFGPTGVGTAKALAEQFDENLLVRIIDMSYMEQHSRLIGAPGVGH 654

658 EGGQLTEAVRRRPVCIIVFEVEKAHVAVFNTLLQVLDGRLTDGQRTYDFRNSVLIIM 717

655 DEGGALTEAIRRPVQVVLFEIEKAHPDVFNVLLQVLDGRLTDGQRTYDFRNTLLIIM 714

718 TSNLGAHLHLAGLTKVTVMEVAROCVREVRKHFPELNLRLDEIVFPDPLSHDLKAKVA 777

715 TSNLGAHYLASQEDGE-DVEAVRPMVMTVRGHPFPEFLNRIIDEIILPKLSRNMGDIV 773

778 RLQKDVAVRLAERGVAVALTDAALDYILAESDYPYCARPIRMWEKVTLSKVVVR 837

774 RLQORVEKLADRRMALADAEALNWLADKGDYDPVYGARPLKRVIQELVDPIAKKLLA 833

838 EBDISNSTVYIDAGAGDLVY 857

Db 834 GEIEDGGVIAVGVTDGQLSF 853

RESULT 12

AD2441

endopeptidase Clp ATP-binding chain B [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AD2441

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AD2441

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-835 <KUR>

A;Cross-references: GB:BA000019; PIDN:BA876783.1; PID:g17134222; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: clpB

C;Superfamily: endopeptidase Clp ATP-binding chain

Query Match 48.8%; Score 2218.5; DB 2; Length 835;

Best Local Similarity 52.6%; Pred. No. 5.7e-91;

Matches 430; Conservative 181; Mismatches 185; Indels 21; Gaps 7;

QY 43 TGIFPQAISSAGGENAAQSAERVINQALKLPSQSPDDIPASSSLIKVIRDAQAQKS 102

Db 12 SGILTKA-----GVNLQKISDRT-EQYIQRPQVSGNSTSVYLGRSLDILLDRAEAHRKD 65

QY 103 RGTHLAVDOLINGLEDSDQI-RDILNVEGVATARVKEVEKLGKGVESASGDTNF 161

Db 66 FQDEYISIEHLHLLAYPKDDRFKGLFQEPALDESKLNIQVRSQ--TVTDQPEGY 123

QY 162 QALKTYGRDIVEQA--GKLDVPIGRDEEIRRVVRIISRRTPKNNPVLIGEPGVGTAVVEG 219

Db 124 QSLKCYGRDLTEAARKGQDPVIGRDRDEIRRTIQLSRRTKNNPVLIGEPGVGTAVVEG 183

QY 220 LAQRIYKGVDPNSLTDVRLISLDMGALVAGAKYRGFEERLKVLEVEDAEKVLFD 279

Db 184 LAQRIYKGVDPNSLTDVRLISLDMGALVAGAKYRGFEERLKVLEVEDAEKVLFD 243

QY 280 EHLVLGAGKTEGSMDDAALFKPMLARGQLRCIGATTLEERYKVKDKDAFFRRFOVVV 339

Db 244 EHTVVGAGATQAGMDAGNLLKPMALARGELRCIGATTLEERYKVKDKDAFFRRFOVVV 303

QY 340 AEPSVPTTISILRLGKLEKEGHHGVRIQDRALINAAQSLARYITGRHLPDKAIDLVDKAC 399

Db 304 DQPSVEDTTISILRLGKLEKEGHHGVRIQDRALINAAQSLARYITGRHLPDKAIDLVDKAC 363

QY 400 ANRVVOLDQPEIDNLEERKMOLETELHALEREKQKASVRLIIVRKELDLRLDKLQPL 459

Db 364 ARUKMEITSKPELDEIDRIKIQLEMEKLSQESDAASRELERLEKELADLKQEQLT 423

QY 460 TMKYRKEKRIDEIRLQKREBELMFSLQEAERRYDLARAADLRYG-----AIQVES 512

Db 424 NTQWQSEKQVINKLQSVKEEIDKVNLEIQQAERNVDLNPAAELKYGNLTDLHRRLEATER 483

QY 513 AIAQLSGTSSSENNVMLTENVGPEHIAEVSVRWGTGIPVTRIGONEKERLIGLADRLKRVV 572

Db 484 ELSQTQGTGKS---LLREEVTEADIAETISKTGTPISKLVSESEKELHLEDELHHRV 540

QY 573 GQNVQAVNAVSEALIRSRAGLGRAQPTGSGFLFGPTGVGTAKALAEQFDENLLVRI 632

Db 541 GQDEAVTAVADAIQSRAGLADPNRPTAFVFLGPTGVGTAKALAEQFDENLLVRI 600

QY 633 IDMSYMEQHSRLIGAPGVGHGEGQLTEAVRRRPVCIIVFEVEKAHVAVFNTLL 692

Db 601 IDMSYMEKHAVSRILGAPGVGHGEGQLTEAVRRRPVCIIVFEVEKAHVAVFNTLL 660

693 QVLDGRLTDGQRTVDFRNVIIIMTSLNGAEHLIAGLTGKVTMEVADCVMEVRKHF 752
 661 QILDDGRVTDAGHTVDFKNTIIIMTSLNGAEHLIAGLTGKVTMEVADCVMEVRKHF 720
 753 PELNRLDEIVVEFPLSHDLQKVKARLQMDVAVRLAERGVALAVTDAAIDYIIAESYDP 812
 721 PEFNRLDEIVVEFPLSHDLQKVKARLQMDVAVRLAERGVALAVTDAAIDYIIAESYDP 780
 813 VYGARPIRMMEKKVVTLEKVVVEEIDENSTVVID 849
 781 VFGARPLKRAIARELETOIAKAILRGFRNDGDTIFVD 817

UTL 13
 905
 opeptidase Clp (EC 3.4.21.-) ATP-binding chain clpB [similarity] - Trypanosoma brucei
 lternate names: ATP-dependent Clp proteinase regulatory chain B
 ontains: adenosinetriphosphatase (EC 3.6.1.3)
 ecies: Trypanosoma brucei
 ccession: E35905
 ate: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 19-Jan-2001
 utesman, S.; Squires, C.; Pichersky, E.; Carrington, M.; Hobbs, M.; Mattick, J.S.; D
 .S. Natl. Acad. Sci. U.S.A. 87, 3513-3517, 1990
 itle: Conservation of the regulatory subunit for the Clp ATP-dependent protease in pr
 ference number: A35905; MUID:90233044; PMID:2185473
 ccession: E35905
 olecule type: DNA
 ssidues: 1-862 <GOT>
 cross-references: GB:M32225
 netics:
 ne: clpB
 nction:
 scription: allows clpP to hydrolyze polypeptides and proteins, probably by a chaperon
 itivity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller
 perfamily: endopeptidase Clp ATP-binding chain
 ywords: ATP; hydrolase; molecular chaperone; nucleotide binding, P-loop; serine prot
 78-205/Region: nucleotide-binding motif A (P-loop)
 6-270/Region: nucleotide-binding motif B
 7-77-604/Region: nucleotide-binding motif A (P-loop)
 5-669/Region: nucleotide-binding motif B
 4/Binding site: ATP (Lys) #status predicted
 3/Binding site: ATP (Lys) #status predicted

Query Match 48.6%; Score 2212.5; DB 1; Length 862;
 Best Local Similarity 53.5%; Pred. No. 1.1e-90;
 Matches 466; Conservative 135; Mismatches 231; Indels 39; Gaps 10;
 7 THKNTETATATHELVAVNAGHAQFTPLHLAALISDPTGIPFOAISAGGENAQAQSAERYI 66
 9 TNAQAQLSDAVESARKHNGFVDPALHVLVFNEDGL-----ASRVL 52
 67 NQ-----ALKKLPSQSPPPDDIPASS-SLIKVIRRAQAQKSGDTHLAVDQILMG 116
 53 RKLNVLEPLAARVAGALPEQRPFRSITFSSDGGCAQHRAENRV--GSLVAVDHLIG 110
 117 LLEDQIRDLNVEGVATARKVEKLRGKGVESAGDTNFQALTKYIGRDLVQEA- 175
 111 LFE---VEALIMAAHASKKAEGALLLELR--KGGKVTSEFQENYQALEKYATDLCKLAE 165
 176 -GKLDPIVGRDEIRVRLTSLRTKNPNVLCEPGVGKTAVVEGLAQRIVKGDVNSLT 234
 166 EGKLDPIVGRDEIRVRLTSLRTKNPNVLCEPGVGKTAVVEGLAQRIVKGDVNSLT 225
 235 DVRLISLDGALVAGAKYRGEFEERLKSVLKVEYDAEGKVLFDIEHLVLGAKGTGSM 294
 226 NTRFLSLDLGALLAGSLRGEFEERLKSVLNVEKSSNGVILFDIEHLVLGAKGTGSM 285
 295 DAANLKPMLARQLCIGATTILEEYRKYVEKDAAFERRPQQVYVAPSPDPTISIRGL 354
 286 DAANLKPMLARQLCIGATTILEEYRKYVEKDAAFERRPQQVYVAPSPDPTISIRGL 345
 355 KEKVEGHGVRIQDRLINAAQLSARVITGRHLDPDKAIDLVEACANVOLDSPQEEID 414

Db 346 KORYEAHHGVQITDNAVVAQAANRYITNRFWPDKAIDLIDEACANRVQLSSRPEAD 405
 QY 415 NLERKEMOLEIELHALEREKDKASKARLIEVEKELDDLRKLOPLTMKRYRKEKERIDEIR 474
 Db 406 ILERKQKQLEIEAKALERDEKASRELRKLVRADQORVEELQPLVSKYNDREORIDELO 465
 QY 475 RLKQKREELMFSLQEAERRYDLARAADLYRGAIOEVESAIAOL-EGTSEENVMILTENVG 533
 Db 466 EMQSRLEDEK-KLERAVRDKMDLAADLQYVPILODRISLKEDIERQKATLVQEKVT 524
 QY 534 PEHIAVVSRTGIPVTRLGQNEKERLIGLARLHKRVVGQNOAVNAVSEAILRSRAGLG 593
 Db 525 EGDVAAVAVRTGIPVVKLSQTRERLLNLSMLHRRYKVGQDEAVERVADAILRAGLIS 584
 QY 594 RAQPTGSLFLGPTGVGTGTELAALAEOLFDENLVRIIDMSEYMEQHSVSRILGAPPG 653
 Db 585 RNSPTASFLFLGPTGVGTGTELVKAAELFDEKHMVRIDMSEYMEQHSVSRILGAPPG 644
 QY 654 YVHBEGGQLTEAVRRPYCVILFDEVEKAVHVAVENTLLOVDDGRLTDGQRTVDFRNS 713
 Db 645 YIGHDEGGQLTEPVRPHAVVLFDEVEKAVHVAVENTLLOVDDGRLTDGQRTVDFRNS 704
 QY 714 VIMTSNLGAELHLAGLTGKVTMEVARCVMEKRPPELLNRLDRIVVDPLSHDOL 773
 Db 705 IIVMTSLNGSEHLNPEETNESYEVLRENVLAAVRSYRPELINELRDDIIVFERRLTEDL 764
 QY 774 RKVARLQMDVAVRLAERGVALAVTDAAIDYIIAESYDPVYGARPIRMMEKKVVTLESK 833
 Db 765 RGVNDNLGAVNRLEKSGSPVLLDDGVKDFILEGHGDANMGARPLRRIEKNIVTEIGR 824
 QY 834 MYVREEIDENST--VYIDAGAGDLVYRVESG 862
 Db 825 MLIAKELPNSTLRVSLPEGGNKLTFGVKRG 855

RESULT 14
 AI0831
 ClpB protein (heat shock protein f84.1) [imported] - Salmonella enterica subsp. enterica
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AI0831
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 th, T.; Comerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
 A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AI0831
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-857 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD05840.1; PID:g16503815; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY2849
 C:Superfamily: endopeptidase Clp ATP-binding chain

Query Match 48.4%; Score 2204; DB 2; Length 857;
 Best Local Similarity 52.1%; Pred. No. 2.6e-90;
 Matches 449; Conservative 155; Mismatches 244; Indels 14; Gaps 9;
 QY 1 MNPEKFTHTMTETATATHELVAVNAGHAQFTPLHLAALISDPTGIPFOAISAGGENAQAQ 60
 Db 1 MLDELTKFQALADQAQLAGHNDQFIPLHLSALINQEGSIRPLITSA-GINAGQ 59
 QY 61 SAERVINQALKKLSQSPPPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQILMGLED 120
 Db 60 -LRTAQALSRLPQVEGTGVDQVPSSELVRVNLCDLQAKRGDNFISSELFVLALES 118
 QY 121 -SQIRDLNVEGVATARKVEKLRGKGVESAGDTNFQALTKYIGRDLVQEA--GK 177
 Db 119 RGTLDLLKSAGATANTITQAEIQMRG--GESVNDQGAEDQKQKKTVDLTERAPQK 176

178 LDPVIGDEEIRRVIRLRSRTKNNPVLIGEPGVGKTAIVVEGLAQRIVKGDVPSNLTVDV 237
 177 LDPVIGDEEIRRVIRLRSRTKNNPVLIGEPGVGKTAIVVEGLAQRIVKGDVPSNLTVDV 236
 238 LISLDMGALVAGAKYRGEPEERLKSVLKEVEDAEGKVLIDFIDHILVLAGKTEGSDMA 297
 237 VLALDMGALVAGAKYRGEPEERLKSVLKEVEDAEGKVLIDFIDHILVLAGKTEGSDMA 296
 298 NLKPKMLARQLRCIGATTLEERKVKYKEDAAEFRRFOQVYVAEPSPVDTTISIRGLKEK 357
 297 NMLKPALARGELHCVGATTLDEVRKVEKDAALERFQKVFVAEPSPVDTTIALRGLKER 356
 358 YEGHGVRIODRALINAAQLSARYITGRHLDPKADLVDEACANVRVQLDSOPEEIDNL 417
 357 YELHHVQITDPAIVAAATLSHRYIADRLQPKADLVDEACANVRVQLDSOPEEIDNL 416
 418 RKMOLIEIHLALEREKDKASKARLLEVRKELDLRDKLPKLTMYKKEKERIDEIRRLK 477
 417 RRIQLKLEQALMKESDEASKRRLDMLEEDDKERQYSELEEEKAKASLSGTQTIK 476
 478 QKREBELMFSIQEARRRYDLARAADLRYGATQEVESAIAQLE-GTSSEENVN--LTENVGP 534
 477 AELEQAKIAIEQARRVGDLMARSELQYKIPLEK--QLEAATQSEKTMRLRNKVTVD 533
 535 EHTAEVSVRWGTGIPVTRLGQNEKERLIGLADRLHKRVGQONQANVAVSAIRLSRAGLR 594
 534 AETAELVARTGIPVSRMLGEGEKRLMRQELHSRVIGQNEAVEAVSNARRSAGLSD 593
 595 AOPTGSFLFGPTGVGKTELAKALAEQLFDDENLVRIDMSYMEHSHSVRLIGAPPCY 554
 594 PNPFGSFLFGPTGVGKTELAKALAEQLFDDENLVRIDMSYMEHSHSVRLIGAPPCY 553
 655 VGEHGGQTEAVRRPYPYCVILFDEVEKAVAVFNTLLQVLDGRLTDGQRTVDPRNSV 714
 654 VGYEGGYLTAVRRPYPYCVILFDEVEKAVAVFNTLLQVLDGRLTDGQRTVDPRNSV 713
 715 IIMTSLGAEHLHAGLTGKVTMEVARDVCMREVRKHFPELLNRLDEIVFDPDLSHQLR 774
 714 VIMTSLGSD-LIQERFGLDYGKMEKMLVGVVSNQFRPEFINRIDEVVFVFLGEGHIA 772
 775 KVARLQMKDVAVRLAERGVALVTDAAALDYLAESYDVPYGPAPIRRMWKKVVTLSK 834
 773 SIAQIQOLQRYKLEBERGYEIHISDALKLLSANGYDVPYGPAPIRRMWKKVVTLSK 832
 835 VREIDENSTVIDAGADLV 856
 833 ILSGELVPGKVRLEANDRRIV 854

SUBT 15
 3070

Species: Agrobacterium tumefaciens
 Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 Accession: AC3070

Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 Age, G.; Giller, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 Karp, P.; Romero, P.; Zhang, S.
 ence 294, 2317-2323, 2001
 Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 er, E.W.

Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 Reference number: AB2577; MUID:21608550; PMID:11743193
 Accession: AC3070
 Status: preliminary
 Molecule type: DNA
 Residues: 1-874 <KUR>
 Cross-references: GB:AE008689; PID:AA44977.1; PID:g17742634; GSPDB:GN00187
 Experimental source: strain C58 (Dupont)
 Genetics:
 Gene: clpB
 Map position: linear chromosome

C:Superfamily: endopeptidase Clp ATP-binding chain
 Query Match 48.3%; Score 2199; DB 2; Length 874;
 Best Local Similarity 51.7%; Pred. No. 4.e-90;
 Matches 447; Conservative 155; Mismatches 253; Indels 10; Gaps 6;

QY 1 MNPEKFTHTKNTTATIAHELAHVAGHAQFTPLHLAGALISDPTGIFFQAISAGGENAAQ 60
 DB 1 MNIDKYSERVGFQSAQTFALAEHQFSEHVLKVLDDQEQWASLIERAGD--AK 58
 QY 61 SAERVINOALKKLSQSPDDIPASSSLIKVIRRAQAOKSGRDTHLAVDQIMGLLED 120
 DB 59 EARLANDAALAKLPKVGSGNGLSLTAPLAKVFTARDLAKAGDSFVTVERLLQALAE 118
 QY 121 SQIRD--LLNEVGVATARVKEVEKLGKGEKGVESAGDTNFQALATYGRDIVEQA--G 176
 DB 119 SSASTSASLKAGATAQALNQVINDIR--KQRTADSANAQGFQALAKYARDLTERAREG 176
 QY 177 KLDPVIGDEEIRRVIRLRSRTKNNPVLIGEPGVGKTAIVVEGLAQRIVKGDVPSNLTVD 236
 DB 177 RLDPVIGRDEIRRTIQVLSRRTKNNPVLIGEPGVGKTAIVVEGLAQRIVKGDVPSNLTVD 236
 QY 237 RLISIDMGALVAGAKYRGEPEERLKSVLKEVEDAEGKVLIDFIDHILVLAGKTEGSDMA 296
 DB 237 KLMLDMGALVAGAKYRGEPEERLKSVLKEVEDAEGKVLIDFIDHILVLAGKTEGSDMA 296
 QY 297 ANLFPKMLARQLRCIGATTLEERKVKYKEDAAEFRRFOQVYVAEPSPVDTTISIRGLKE 356
 DB 297 SNLAKPALARGELHCVGATTLDEVRKVEKDAALERFQKVFVAEPSPVDTTISIRGLKE 356
 QY 357 YEGHGVRIODRALINAAQLSARYITGRHLDPKADLVDEACANVRVQLDSOPEEIDNL 416
 DB 357 KYEGRHKVIRISDSALVAAATLSNRYITDRFLPKADLVDEACANVRVQLDSOPEEIDNL 416
 QY 417 ERKMOLEIHLALEREKDKASKARLLEVRKELDLRDKLPKLTMYKKEKERIDEIRRL 476
 DB 417 DRIIQLKIEREALQKTDQSSVDRLKLEDEADTBEKADALTARWQAEKQKLGHADL 476
 QY 477 KOKREELMFSIQEARRRYDLARAADLRYGATQEVESAIAQLEG-TSSEENVNLTENVGPE 535
 DB 477 KKRLEDEARNELAIQRNGQFQFAGELTYGIIPLGKELAAEAARDSSGAGSMQVQVETPD 536
 QY 536 HIAVSVRWGTGIPVTRLGQNEKERLIGLADRLHKRVGQONQANVAVSAIRLSRAGLR 595
 DB 537 NIAHVSVRWGTGIPVTRLGQNEKERLIGLADRLHKRVGQONQANVAVSAIRLSRAGLR 596
 QY 596 QOPTGSFLFGPTGVGKTELAKALAEQLFDDENLVRIDMSYMEHSHSVRLIGAPPCY 655
 DB 597 NRPIGSIFFIGPTGVGKTELAKALAEQLFDDENLVRIDMSYMEHSHSVRLIGAPPCY 656
 QY 656 GHEGGQTEAVRRPYPYCVILFDEVEKAVAVFNTLLQVLDGRLTDGQRTVDPRNSV 715
 DB 657 GYEGGQTEAVRRPYPYCVILFDEVEKAVAVFNTLLQVLDGRLTDGQRTVDPRNSV 716
 QY 716 IIMTSLGAEHLHAGLTGKVTMEVARDVCMREVRKHFPELLNRLDEIVFDPDLSHQLR 775
 DB 717 IIMTSLGSE-FMTQMGNDNDVDSVRELVMERVSRHPEFINRIDDIILFHLRERDEMGA 775
 QY 776 VARLQMKDVAVRLAERGVALVTDAAALDYLAESYDVPYGPAPIRRMWKKVVTLSK 835
 DB 776 IVEIQLKRLVSLADRKITLEDEADSWLANQYDPAVGPAPIRRMWKKVVTLSK 835
 QY 836 VREIDENSTVIDAGADLVYRVE 860
 DB 836 LGGEIPDGSRVKVTSGTDRLLFKVK 860

Search completed: February 13, 2004, 01:34:13
 Job time : 51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 13, 2004, 00:52:50 ; Search time 39 Seconds
(without alignments)
1098.496 Million cell updates/sec

le: US-09-812-350-17

fect score: 4550

uence: 1 MNPKEFKTKNETTATAHEL.....VKMKRIEIEDDDNEEMIED 911

ring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

urched: 127863 seqs, 47026705 residues

al number of hits satisfying chosen parameters: 127863

imum DB seq length: 0

imum DB seq length: 2000000000

-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

abase : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ult No.	Query Match	Score	Length	DB ID	Description
1	4545	99.9	911	1	H101 ARATH
2	2288	50.3	872	1	CLPB SYN3
3	2284	48.9	868	1	CLP TRVB
4	2224	48.9	883	1	CLPB SYN7
5	2215.5	48.7	854	1	CLPB THETH
6	2193	48.2	857	1	CLPB ECOLI
7	2162.5	47.5	856	1	CLPB HASIN
8	2158.5	47.4	848	1	CLPB MYCTU
9	2145	47.1	852	1	CLPB CORGL
10	2104	46.2	878	1	CLPB TRSPA
11	2039	44.8	860	1	CLPB BACNO
12	1890	41.5	908	1	H104 YEAST
13	1884.5	41.4	856	1	CLPB HELPJ
14	1876.5	41.2	856	1	CLPB HELPY
15	1844.5	40.5	812	1	HS78 CANAL
16	1805.5	39.7	811	1	HS77 YEAST
17	1759.5	38.7	810	1	CLPB BACSU
18	1759.5	38.7	848	1	CLPB MYCTU
19	1752.5	38.5	848	1	CLPB MYCTU
20	1737.5	38.2	819	1	CLPC GULTH
21	1735.5	38.1	922	1	CLPA PEA
22	1733.5	38.1	854	1	CLPC CYACA
23	1731	38.0	923	1	CLAB LYCES
24	1703.5	37.4	926	1	CLAA LYCES
25	1697.5	37.3	821	1	CLPC PORPU
26	1658	36.4	714	1	CLPB MYCPE
27	1655	36.4	715	1	CLPB MYCPN
28	1654	36.4	874	1	CLPA BRANA
29	1639.5	36.0	828	1	HLVB TREHY
30	1613.5	35.5	870	1	CLPC CHLMU
31	1602	35.2	845	1	CLPC CHLPN
32	1602	35.2	854	1	CLPC CHLTR
33	1547.5	34.0	885	1	CLPC_ODOSI

34	1478.5	32.5	945	1	ERD1 ARATH
35	1422	31.3	748	1	CLPE LACLC
36	1418	31.2	748	1	CLPE LACLA
37	1396.5	30.7	752	1	CLPE_STRPN
38	1331.5	29.3	758	1	CLPA_ECOLI
39	1299.5	28.6	793	1	CLPA_RHOBL
40	1256	27.6	763	1	CLPL LACLA
41	534.5	11.7	677	1	SKD3 MOUSE
42	534.5	11.7	707	1	SKD3_HUMAN
43	525.5	11.5	677	1	SKD3_RAT
44	463.5	10.2	351	1	ADPR_RHOER
45	421	9.3	371	1	AMIB_PSEAE

RESULT 1

H101 ARATH	STANDARD;	PRT;	911 AA.
AC	P42730; Q8WAF2; Q9LE57;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Heat shock protein 101.		
GN	HSP101 OR AT1G74310 OR F1017.2.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. Columbia; TISSUE=Leaf;		
RX	MEDLINE=95170291; PubMed=7866032;		
RA	Schirmer E.C., Lindquist S., Vierling E.;		
RT	"An Arabidopsis heat shock protein complements a thermotolerance defect in yeast."		
RL	Plant Cell 6:1899-1909(1994).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. Columbia;		
RX	MEDLINE=20226114; PubMed=10760305;		
RA	Hong S.-W., Vierling E.;		
RT	"Mutants of Arabidopsis thaliana defective in the acquisition of tolerance to high temperature stress."		
RL	Proc. Natl. Acad. Sci. U.S.A. 97:4392-4397(2000).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. Columbia;		
RX	MEDLINE=21016719; PubMed=11130712;		
RA	Theologis A., Ecker J.R., Palm C.J., Federapiei N.A., Kaul S., White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y., Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A., Militischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I., Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D., Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;		
RT	"Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana."		
RL	Nature 408:816-820(2000).		
RN	[4]		
RP	SEQUENCE OF 1-460 FROM N.A.		
RC	STRAIN=cv. Columbia;		

Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 SSP consortium (Salk/Stanford/PCEC)";
 Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THERMOTOLERANCE.
 -!- INDUCTION: By heat shock.
 -!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; U13949; AAA67927.1; --
 EMBL; AF218796; AAF26423.1; --
 EMBL; AC020579; AAG52410.1; --
 EMBL; AY062596; AAL32674.1; --
 PIR; F96771; F96771.
 InterPro; IPR003593; AAA ATPase.
 InterPro; IPR001270; Chaprin-clpa/B.
 InterPro; IPR004176; Clp_N.
 Pfam; PF000004; AAA; 2.
 Pfam; PF02861; Clp_N; 2.
 SMART; SM00300; CLPPTREASEA.
 SMART; SM00382; AAA; 2.
 PROSITE; PS00870; CLPB_1; 1.
 PROSITE; PS00871; CLPB_2; 1.
 Chapterone; Heat shock; ATP-binding; Repeat.
 DOMAIN 164 410
 I.
 NP_BIND 532 723
 II.
 NP_BIND 207 214
 ATP (POTENTIAL).
 NP_BIND 606 613
 ATP (POTENTIAL).
 CONFLICT 141 141
 V -> F (IN REF. 4).
 CONFLICT 595 595
 P -> A (IN REF. 1).
 SEQUENCE 911 AA; 101294 MW; 191EC1853B0C4CB9 CRC64;

Query Match 99.9%; Score 4545; DB 1; Length 911;

Best Local Similarity 99.9%; Pred. No. 3.2e-191; Matches 910; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MNPEFTHKNTETATHEAVNAGHAQFTPLHLAGALISDPTGIFPQAISSAGGENAAQ 60
 1 MNPEFTHKNTETATHEAVNAGHAQFTPLHLAGALISDPTGIFPQAISSAGGENAAQ 60
 61 SAERVINQALKLPQSPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLIMGLLED 120
 61 SAERVINQALKLPQSPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLIMGLLED 120
 121 SQIRDLNNEVGAVATRVKSEVEKLKRGKGVESAGDTNFQALKTYGRDLVEQAGKLDLP 180
 121 SQIRDLNNEVGAVATRVKSEVEKLKRGKGVESAGDTNFQALKTYGRDLVEQAGKLDLP 180
 181 VTGRDEIRRVRLISRRTPKNNPVLIGEPGVKTAVVEGLAQRIKVGDPVNSLTDVRLIS 240
 181 VTGRDEIRRVRLISRRTPKNNPVLIGEPGVKTAVVEGLAQRIKVGDPVNSLTDVRLIS 240
 241 LDMGALVAGAKYRGEFEERLKSVLKEVEDAEKGVLFIDEHLVLAGKTEGSDMAANLF 300
 241 LDMGALVAGAKYRGEFEERLKSVLKEVEDAEKGVLFIDEHLVLAGKTEGSDMAANLF 300
 301 KPMARGLQRCIGATTLEBYRYKVEKDAFERFQVVAEPSVPTTISILRGLKEKVEG 360
 301 KPMARGLQRCIGATTLEBYRYKVEKDAFERFQVVAEPSVPTTISILRGLKEKVEG 360
 361 HGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDCAANRVQLDSQPEIDNLERK 420
 361 HGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDCAANRVQLDSQPEIDNLERK 420
 421 MLEIELHALEREKDKASKARLIEVRKELDDLRDKLQPLTMKYRKEKERIDEIRLQKQR 480

Db 421 MLEIELHALEREKDKASKARLIEVRKELDDLRDKLQPLTMKYRKEKERIDEIRLQKQR 480
 Qy 481 EELMFSIQEAEERYDYLARAADLYGALQFVESAIQAQLEGTSSEENVMLTENVGPEHIAEV 540
 Db 481 EELMFSIQEAEERYDYLARAADLYGALQFVESAIQAQLEGTSSEENVMLTENVGPEHIAEV 540
 Qy 541 VSRWTGIPVTRLGQNEKERLIGLADRLHKEVVGQNAVNAVSEAILRSRAGLRGQOQPTG 600
 Db 541 VSRWTGIPVTRLGQNEKERLIGLADRLHKEVVGQNAVNAVSEAILRSRAGLRGQOQPTG 600
 Qy 601 SFLPLGTGVTGKTELAKALAEQFDDENLIVRIMDMGYMEQHSVSRLLIGAPPGYVGHHEG 660
 Db 601 SFLPLGTGVTGKTELAKALAEQFDDENLIVRIMDMGYMEQHSVSRLLIGAPPGYVGHHEG 660
 Qy 661 GQLTEAVRRPYCVILFDEVEKAHVAVENTLLOVLDGRLTDGQRTVDPRNSVIIMTSN 720
 Db 661 GQLTEAVRRPYCVILFDEVEKAHVAVENTLLOVLDGRLTDGQRTVDPRNSVIIMTSN 720
 Qy 721 LGAEHLLAGLTGKVTMEVARDCVMRVVRKHFRPELLNRLDEIVVFDPLSHDQLRKVARLQ 780
 Db 721 LGAEHLLAGLTGKVTMEVARDCVMRVVRKHFRPELLNRLDEIVVFDPLSHDQLRKVARLQ 780
 Qy 781 MKOVAVRLAERGVALAVTDAAALDYILAESYDYPYGARPIRRMMKKVVTLSKVVVREEI 840
 Db 781 MKOVAVRLAERGVALAVTDAAALDYILAESYDYPYGARPIRRMMKKVVTLSKVVVREEI 840
 Qy 841 DENSTVYIDAGDLVYRVESGGLVDASTGKSDVLIHTANGPKRSDAQAQVKMKIEEI 900
 Db 841 DENSTVYIDAGDLVYRVESGGLVDASTGKSDVLIHTANGPKRSDAQAQVKMKIEEI 900
 Qy 901 EDDNNEEMIED 911
 Db 901 EDDNNEEMIED 911
 RESULT 2
 CLPB SYN3
 ID CLPB SYN3 STANDARD; PRT; 872 AA.
 AC P74361;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ClpB protein.
 GN CLPB OR SLR1641.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RK MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 Miyajima T., Hiroseawa M., Sugiyura M., Sasamoto S., Kimura T.,
 Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
 Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
 Yamada M., Yasuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -!- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR
 ATP-DEPENDENT PROTEASE (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; D90914; BAA18456.1; --
 PIR; S76197; S76197.


```

RN      SEQUENCE FROM N.A.
RP      MEDLINE=96326327; PubMed=8759846;
RX      Eriksson M.J., Clarke A.K.;
RA      "The heat shock protein ClpB mediates the development of
RT      thermotolerance in the cyanobacterium Synchococcus sp. strain PCC
RT      7942."
RL      J. Bacteriol. 178:4839-4846(1996).
RR      -I- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR
CC      ATP-DEPENDENT PROPEASE (BY SIMILARITY).
CC      -I- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U20646; AAB09631.1; -.
DR      InterPro: IPR003593; AAA_ATPase.
DR      InterPro: IPR003959; AAA_ATPase_centr.
DR      InterPro: IPR001270; Chaprinin_clpA/B.
DR      InterPro: IPR004176; Clp_N.
DR      Pfam: PF00004; AAA; 1.
DR      Pfam: PF02861; Clp_N; 2.
DR      PRINTS: PR00300; CLPPROTEASEA.
DR      SMART: SM00382; AAA; 2.
DR      PROSITE: PS00870; CLPAB_1; 1.
DR      PROSITE: PS00871; CLPAB_2; 1.
KW      Chaperone; ATP-binding; Repeat.
FT      DOMAIN 172 420
FT      I.
FT      DOMAIN 546 737
FT      II.
FT      NP_BIND 217 224
FT      NP_BIND 620 627
FT      ATP (POTENTIAL).
FT      ATP (POTENTIAL).
SQ      SEQUENCE 883 AA; 99594 MW; 82E0F0FF70A0A49 CRC64;

```

Query Match	48.9%	Score 2224	DB 1	Length 883
Best local similarity	51.3%	Pred. No. 4.3e-90		
Matches	444	Conservative 186	Mismatches 220	Indels 16
Gaps	7			
QY	2	NPEKFTKNTETATAEHALAVAGHAQFTPLHLAGALISDPTGIFPQAISSAGGENA--A	59	
DB	14	NNNQFTKEAWELVRTDVAQQAQHQIIESEHLFTALQEP-GLALNLIKAGLEAARQLQ	72	
QY	60	QSAERVINOALKKLPSQSPPPDDIPASSSLIKVIRPAAQAQKSRGTHLAVDOLIMGLE	119	
DB	73	QFTERFI---ARQPKVSGNSVYLGRSLDQLLDQADQFRKDFGDEFISVEHLIILSFR	128	
QY	120	DSQIRDLN-EVGVATARYKSEVEKLGRGEGKKVESASGDTNFQALKTKYGRDLVEQA--G	176	
DB	129	DSRFGRLLSQEPKVDKQRLRIIQIRGSG--KVTDONPEGKVEALEKYGRDLTEMARRG	186	
QY	177	KLDPVIGRDEETRRVRIISRRTPKNPNVLIGBPGVKTAVVEGLAQRIKGVGVNPSLTDV	236	
DB	187	KLDPVIGRDEETRRITQIISRRTPKNPNVLIGBPGVKTAIEGLAQRIINGDVVPOKLDK	246	
QY	237	RLISLDMGALVAKAYRGPFERELKSVLKEVEDAEKGVILFTDEIHLVLGAGKTEGMDA	296	
DB	247	RLIALDNGALIAGKFGPFERELTAVLKEVTDSEGIILFIDEMTIVVAGAVQGSMDA	306	
QY	297	ANLFKPNLARGQLRCIGATTLLBEYRKVKYBKDAAFERRFOOVYVAEPSVPTTISILGLKE	356	
DB	307	GNNLKTMLARGELRCIGATTLLGKRYIIEKDAALERFQOVFDQPTVETDTISILGLKE	366	
QY	357	KYEGHGVRIQDRALINAAQLSARYITGTGHLDPKADLDLVEACANVRVOLDSPSEIDNL	416	
DB	367	RVEVHGVRISDNLVAAVAVLSTRYISDRFLDPKADLDLVDENARLKMITSKPELDEI	426	
QY	417	ERKMQLEIETHALEREKDKASKARLIEYRKLEDDLDRLDKLOPLTMKYRKEKERIDEIRL	476	
DB	427	DRKIQLEMERLSLOKESLASOERLRIEKELEADIKEBQSRSLSSQOAKEDVITIQSV	486	

ULT 4
 B_SYN07
 CLPB SYN07 STANDARD; PRT; 883 AA.
 P53533;
 01-OCT-1996 (Rel. 34, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-NOV-1997 (Rel. 35, Last annotation update)
 CLPB protein.
 CLPB.
Synechococcus sp. (strain PCC 7942) (*Anacystis nidulans* R2).
 Bacteria; Cyanobacteria; Chroococcales; *Synechococcus*.
 NCBI TaxID=1140;

706 LTNLSGLSPLEGLQKGMVEYRDEVRKVLQHPFELNRLDRIIVFPLTKRQIRQI 765
777 ARLOKQDVAVRLAERGVAVLADVAALDYLAESYDPVVGARPIREWMEKKVVTLSKMWV 836
766 VEQISYLRARLAERKRISLLELTAERKFLAERGYPVVGARPIRVIQRELETPLAQKIL 825
837 REEIDENSTVYIDAGAGDLVYRV 859
826 AGEVKEGDRVQVDVGPAGLVFVAV 848

JUL 6

CLPB ECOLI STANDARD; PRT; 857 AA.
P03615;
21-JUL-1986 (Rel. 01, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
15-SEP-2003 (Rel. 42, Last annotation update)
ClpB protein (Heat shock protein F84.1).
CLPB OR HTPM OR B2592 OR C3114 OR Z3886 OR ECS3455.
Escherichia coli,
Escherichia coli O6, and
Escherichia coli O157:H7.
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
NCBI TaxID=562, 217992, 83334;
[1]
SEQUENCE FROM N.A.
MEDLINE=90239044; PubMed=2185473;
Gottesman S., Squires C., Pichersky E., Carrington M., Hobbs M.,
Mattick J.S., Dairymple B., Kuramitsu H., Shiroza T., Foster T.,
Clark W.P., Ross B., Squires C.L., Maurizi M.R.;
"Conservation of the regulatory subunit for the Clp ATP-dependent
protease in prokaryotes and eukaryotes."
Proc. Natl. Acad. Sci. U.S.A. 87:3513-3517 (1990).
[2]
SEQUENCE FROM N.A.
STRAIN=K12 / MG1655;
MEDLINE=97426617; PubMed=9278503;
Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
"The complete genome sequence of Escherichia coli K-12."
Science 277:1453-1474 (1997).
[3]
SEQUENCE FROM N.A.
STRAIN=K12;
MEDLINE=97349980; PubMed=9205837;
Yamamoto Y., Aliba H., Baba T., Hayashi K., Inada T., Isono K.,
Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
Oshima T., Oyama S., Saito N., Saueki G., Satoh Y., Sivasubraman S.,
Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
Yamagata S., Horiuchi T.;
"Construction of a contiguous 874-kb sequence of the Escherichia coli
- K12 genome corresponding to 50.0-68.8 min on the linkage map and
analysis of its sequence features."
DNA Res. 4:91-113 (1997).
[4]
SEQUENCE FROM N.A.
STRAIN=O6:H1 / CFT073 / ATCC 700928;
MEDLINE=22388234; PubMed=12471157;
Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
"Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli."
Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
[5]
SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Rao G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
Apodaca J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,
Welch R.A., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
Nature 409:529-533 (2001).
[6]
SEQUENCE FROM N.A.
STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shingawa H.;
"Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12."
DNA Res. 8:11-22 (2001).
[7]
SEQUENCE OF 1-593 FROM N.A.
STRAIN=K12;
RX MEDLINE=91294165; PubMed=1906060;
Kitagawa M., Wada C., Yoshioka S., Yura T.;
"Expression of ClpB, an analog of the ATP-dependent protease
regulatory subunit in Escherichia coli, is controlled by a heat shock
sigma factor (sigma 32)."
J. Bacteriol. 173:4247-4253 (1991).
[8]
SEQUENCE OF 753-857 FROM N.A.
MEDLINE=82247208; PubMed=6285294;
Shen W.-F., Squires C., Squires C.L.;
"Nucleotide sequence of the rrmG ribosomal RNA promoter region of
Escherichia coli."
Nucleic Acids Res. 10:3303-3313 (1982).
[9]
SEQUENCE OF 1-31 FROM N.A.
STRAIN=K12 / W3110;
RX Ogura T., Tomoyasu T.;
Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
[10]
SEQUENCE OF 1-14; 150-157; 355-364 AND 452-460.
MEDLINE=92062147; PubMed=1953774;
Pontis E., Sun X.Y., Joernvall H., Krook M., Reichard P.;
"ClpB proteins copurify with the anaerobic Escherichia coli
reductase."
Biochem. Biophys. Res. Commun. 180:1222-1226 (1991).
[11]
IDENTIFICATION AS A HEAT SHOCK PROTEIN.
MEDLINE=91294165; PubMed=2066329;
Squires C.L., Pedersen S., Ross B.M., Squires C.;
"ClpB is the Escherichia coli heat shock protein F84.1."
J. Bacteriol. 173:4254-4262 (1991).
-1- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR ATP-
DEPENDENT PROTEASE.
-1- ALTERNATIVE PRODUCTS:
Event=Alternative initiation;
Comment=2 isoforms, ClpB (shown here) and ClpB-3, are produced
by alternative initiation;
-1- INDUCTION: By heat shock.
-1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M29364; AAA24422.1; -

	ENBL; AE000345; AAC75641.1; -	
	ENBL; D90887; BAA16476.1; -	
	ENBL; AE016764; AAN81563.1; ALT_INIT.	
	ENBL; AE005489; AGS7705.1; ALT_INIT.	
	ENBL; AP002562; BAB36878.1; -	
	ENBL; X57620; CAA40846.1; -	
	ENBL; V00350; CAA23639.1; -	
	ENBL; U50134; AAA92959.1; -	
	PIR; G65037; D35905.	
	PIR; G91060; G91060.	
	PDB; 1JBK; 05-JUN-02.	
	PDB; 1KHV; 04-DEC-02.	
	SWISS-2DPAGE; P03815; COLI.	
	ECO2DBASE; E072.0; 6TH EDITION.	
	ECO2DBASE; F084.1; 6TH EDITION.	
	EcoGene; EG10157; clpb.	
	InterPro; IPR003593; AAA_ATPase.	
	InterPro; IPR003959; AAA_NTPase_centre.	
	InterPro; IPR001270; Chaprinin_cipa/B.	
	InterPro; IPR004176; Clp_N.	
	Pfam; PF00004; AAA; 2.	
	Pfam; PF02861; Clp_N; 2.	
	PRINTS; PR00300; CLPPROTEBASEA.	
	SMART; SM00382; AAA; 2.	
	PROSITE; PS00870; CLPB_1; 1.	
	PROSITE; PS00871; CLPBAB_2; 1.	
	Chapone; ATP-binding; Repeat; Heat shock; Alternative initiation;	
	Complete proteome; 3D-structure.	
	CHAIN 1 CLPB PROTEIN, ISOFORM CLPB.	
	INIT_MET 149 857	
	DOMAIN 149 149	
	DOMAIN 161 409	
	I.I. 531 722	
	NP_BIND 206 213	
	NP_BIND 605 612	
	CONFLICT 96 97	
	CONFLICT 122 122	
	SEQUENCE 857 AA; 95585 MW; FD38CD96B2F7C32A CRC64;	
	Very Match 48.2%; Score 2193; DB 1; Length 857;	
	Best Local Similarity 51.7%; Pred.No. g.3e-89;	
	Hits 447; Conservative 154; Mismatches 239; Indels 24; Gaps 9;	
	1 MNPEKTKHNTETATAHELAVNAGHAQTPLHLGALISDPTGIPQAISSAGGENAAQ 60	
	1 MRDLRTNKPOLADLAQSIALGHNDNQFTLEPHLSALLNQEGSVSLTSA-GINAQ 59	
	61 SAERVNOALKKLPSOSPPDDITPASSLLIKVTRRAQAQRSGDTHLVDQLINGLED 120	
	60 -LRDTNQALNRLPQVEGTGGYQSDLVRLNLCDKLAQRKGDNFISSFELVALES 118	
	121 -SOIRDILLEVGATVRARVKSEVLGRKEGKVESAGDTNFOLAKTYGRDVLOEA--GK 177	
	119 RGLADILKAAGATTANTICAIQMKG--GESVNDQGARDQCALKKYTIDITERAQEK 176	
	178 LDPVIGRDEIRRVRIKRRTKNPNVLIGEPCGVKTAVVEGLAQIVKGDVPNSLTDR 237	
	177 LDPVICRDEIRRTIQVLRRTKNPNVLIGEPCGVKTAVEGLAQRIINGEVEGLKGR 236	
	238 LISLDMGALVAGKYGEFERLKSVLKEVEDBGKVIIFIDEIHVLVLAGTGESMDAA 297	
	237 VLALDMGALVAGKYGEFERLKSVLNDAKOEGNVILFIDELHTMVAGKADGMADAG 296	
	298 NLFKPMLARGCLRCIGATTLLEEYRKVKVEKDAAERRFRQQVYAEPSPVTISTLRGLEK 357	
	297 NLKPALARGELCHCGATTIDETRQIEKDALARFPQKVFPAEPSVESDIALLRGELKER 356	
	358 YEGHHGVRIODRALINAQAASARYITGRHLPDKAIDLVBECANVRVQDSOPFEEDNLE 417	
	357 YELHHVQIIDPAIVAATASHYIADRQLPKAKAIDILDEASSINQWIDSXPEELDRLD 416	
	418 KRMQLEIEUHALERKOKASKARLLIEVRKELDDLDRDKLQPLTMKYRKKEKERIDEIRRK 477	

PIR; F64098; F64098.
TIGR; H10859; -.
InterPro; IPR003593; AAA ATPase.
InterPro; IPR003959; AAA ATPase centr.
InterPro; IPR001270; Chaprinin_clpA/B.
InterPro; IPR004176; Clp_N.
Pfam; PF00004; AAA; 2.
Pfam; PF02861; Clp_N; 2.
PRINTS; P00300; CLPPTASEA.
SMART; SM00382; AAA; 2.
PROSITE; PS00870; CLPAB_1; 1.
PROSITE; PS00871; CLPAB_2; 1.
Chaperone; ATP-binding; Repeat; Complete proteome.
DOMAIN 161 409 I.
DOMAIN 531 722 II.
NP_BIND 206 213 ATP (POTENTIAL).
NP_BIND 605 612 ATP (POTENTIAL).
SEQUENCE 856 AA; 95837 MW; B9958ED9D03EA3B CRC64;
Query Match 47.5%; Score 2162.5; DB 1; Length 856;
est Local Similarity 52.1%; Pred. No. 2e-87;
atches 444; Conservative 153; Mismatches 245; Indels 11; Gaps 8;
1 MNPEKFTKTTTATAHNAVAGHQAFTPLHLAALISDPTG-IPQAISSAGGENAA 59
1 MNIEKFTTKFOEALUSEROSLAGKNDQFIEPVHLLTALLNQQGSIAP--ILTASGVNVA 58
60 QSAERVINQALKIPSOQPPDDIPASSSLIKVIRRAQAQAQKSRGDTHLAVDQIMGLLE 119
59 -LLRNELKTELKLPQVINGGVDQVLSQLNLLNLCUKFAQQNQDFISELFLFAALE 117
120 D-SQIRDLLNEGVATARKVEKRLKRGKVKVESAGDTNFQALTYGRDLVQEA--G 176
118 ERGTISDILKCCGAKKEQISOAIHQIRG--QNVNDQNAEBSROALEKYITDILTAESG 175
177 KLDPIVGHDEIRVRLISRTKNNPVLGEPGVGTATVVEGLAQRIKVGDPVNSITDV 236
176 KLDPIVGHDEIRAIQVLRKNNPVLGEPGVGTATVVEGLAQRIKVGDPVNSITDV 235
237 RLISLDGALVAGAKYGEFERLKSILKEVEDAEGKVLFIIDEHLVLGAGKTEGSMDA 296
236 RVLSDMGALVAGAKYGEFERLKAIVNELSKKEGRVILFIDEIHTVMGAGKTDGMDA 295
297 ANLKPMLARQLKICGATTLVEKRVKVEKDAEERFPQVYVAEVPVPTITSLRLKE 356
296 GNLLKPSLARGELCHVGATTTDEYRQVIEKDAALERRFKVFDVPEPVEDTIALRLKE 355
357 KYEGHGVRIODRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQDSQPEEDNL 416
356 RYEHVHVDTIDPAIVAAATLSHYISDRQPDKAIDLIDEAASSIRMEIDSKPEPDLRL 415
417 ERKRMQLEIEHUALEREKOKASKARLIEVRKELDDLRLKQLPLTKYRKERIDEIRRL 476
416 ERRIITQLKEQALQKEDEASRKLMELEKAEKEREVAELEEVKSEKATLSGQHI 475
477 KQKEELMFSIQEAERRYDLARADLYALGATQVESAIALEGTSSENWMLTVNGPEH 536
476 KQELDTAKTEQARRAGDIAKMSLEQVGRIPDLKQLEQAETSEGKMTLLRYVTDEE 535
537 IAEVSVRWGTPVTRLGQNEKRLIGLADRLHKRVGNQAVNSAIIKRSRAGLGRAQ 596
536 IAEVLSKATGIPVSKMGEKEKRLMEDELUKRVIGQEAADVAVANIRRSRAGLSDPN 595
597 QPTGSFLFGPTGVGKTEKALAKAEQLFDDENLLVRIDMSYMEQHSVSRLLIAGPPGVG 656
596 RPIGSFLFGPTGVGKTEKCTLAKFLFDSADVAMVDMSEMEKHSVSRVLGAPPGV 655
657 HEEGQLTAEVRRPVCVILFDEVEKAHVAVNTLLQVLDGRITDCGRTVDFRNSVII 716
656 YEEGGLTAEVRRPYSVILLDEVEKAHADVFNILLQVLDGRITDCGRTVDFRNTVII 715
717 MTSNLGASHLAGLTGKVTMEVARDVCMVREVRKFRPELLNRLDEIVVDFPLSHDQLRV 776

Db 716 MTSNLGSD-LIQGNKDESYSEM-KALVMSVVSQHFRPEFINRIDETVVFHPLGKINIRAI 773
QY 777 ARLOMKDVAVRLAERGVALAVTDAALDYILAESYDPVYGARPIRRWMEKKVVTLSKMVV 836
Db 774 ASIQLERLAKRMETRGYELVFTDALLDFIGVGYDPIYGARLEKRAIQOEIENSLAQILL 833
QY 837 REIDENSTVYID 849
Db 834 SGALLPGKVVTID 846

RESULT 8
CLPB MYCTU
ID CLPB MYCTU STANDARD; PRT; 848 AA.
AC OS3719;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ClpB protein.
GN CLPB OR RV0384C OR MT0397 OR MTV036.19C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
Ravick K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544 [1998].
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bisai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR
CC ATP-DEPENDENT PROTEASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL021931; CAAL7390.1; -.
CC EMBL; AE006944; AAK44619.1; -.
CC PIR; C70834; C70834.
CC TIGR; MT0397; -.
CC TubercuList; RV0384c; -.
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR003959; AAA ATPase centr.
CC InterPro; IPR001270; Chaprinin_clpA/B.
CC InterPro; IPR004176; Clp_N.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF02861; Clp_N; 2.

PRINTS; PRO0300; CLPPTROTEASEA.
SMART; SM00382; AAA; 2.
PROSITE; PS00870; CLPAB 1; 1.
PROSITE; PS00871; CLPAB 2; 1.
Chaperone; ATP-binding; Repeat; Complete proteome.
DOMAIN 161 410 I.
DOMAIN 533 724 II.
NP_BIND 206 213 ATP (POTENTIAL).
NP_BIND 607 614 ATP (POTENTIAL).
SEQUENCE 848 AA; 92568 MW; 41CE8DFA9D5EBAC9 CRC64;

Query Match 47.4%; Score 2158.5; DB 1; Length 848;
Best Local Similarity 52.2%; Pred. No. 2.9e-87;
Matches 447; Conservative 150; Mismatches 227; Indels 33; Gaps 9;

2 NPEKTHKTNETIATAHELVNAGHAQFTPLHAGALISDPTGTFPPQAISAGGENAQAOS 61
5 NP---TYKQALTAALQAASAGNPIRPAHLMAULTQNDGIAAPLLEAVGVEPATVR 61
62 AERVINOALKKLPSSQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLLE-D 120
62 AE--TORLLDRLPQATGASTQPLSRSLAATTAAQQLATLDEYVSTEHVVGATGD 119
121 SQIRDLNVEGVATARKSEVKEKRGKKEKVESASGDTNFOALKTYGRDLVQA--GKL 178
120 SDVAKLTHGASPAQLREAFVKVRG--SARVTSPEPEATYQALQKYSTDLTARAREGKL 177
179 DPVIGRDBEIRRVVRIISRRTPNNPVLIGPGVKTAVVEGLAQRIKVGDPVNSLTDVRL 238
178 DPVIGRDBEIRRVVRIISRRTPNNPVLIGPGVKTAVVEGLAQRIKVGDPVNSLTDVRL 237
239 ISLDMGALVAGAKYRGFEERLKSVLKEVEDAEKVLIFIDEHLVLGAKT--EGSMDAA 237
238 VALDGLSWAGSKYRGFEERLKAVIDLKNASQIITFIDELHTIVGACATGEGAMDAG 237
298 NLFKPMILARGLRCIGATTLEERKYEKEDAAEFERRRQOYVYAPSPVDTISIRLGEK 357
298 NMTKPMILARGLRVGATTLDEYRKHKEDAALERRFPQYVYGEPSVEDTIGLRGLKOR 357
358 YEGHGVRIIDRALINAQISARYITGRHLPDKAIDLVDEACANVRVQLDSQPEEDINLE 417
358 YEYHGVRIIDRALINAQISARYITGRHLPDKAIDLVDEACANVRVQLDSQPEEDINLE 417
418 RKMQLIEHLALEREKOKASKARLIEVRKELDLRDKLOPLTKYRKEKERIDEIRRLK 477
418 RLVRLEIEEMALSKEDESAERLAKRSELADQKEKLAELTRWQNEKNALEIVRDLK 477
478 QKEELMFSIQEARRVDLARAADIRYGAIQVESAI-AQLEGTSSEENVMLTENTYGP 536
478 EQLEALRGESERAEKRGDGLAKAELRYGRIPVEVEKLDLDAALPOAQAQEQVMLKEEYGP 537
537 IAEVSRWTGIPVTRLGQNEKERLIGLADLRHVRVQGNQAVNAVSEAILRSRAGLGRAQ 596
538 IADVSAWTGIPAGRLLEGATKALLRMEDLGRVIGQKAAVAVSDAVERBAGVSDEN 597
597 OPTGSLFLPGTGVGKTELAKALAQOLFDDENLLRVIDMSEYMEQHSVRLIGAPPGYV 656
598 RPTGAFNPLGPTGVGKTELAKALADFLDDERAMVRIDMSEYGEKHTVARLIGAPPGYV 657
657 HESGQQLTEAVRRPVCVILFDEVEKHAFAVNTLLQVLDGRLTGQGTVDVFRNSVII 716
658 YEAGQQLTEAVRRPVCVILFDEVEKHAFAVNTLLQVLDGRLTGQGTVDVFRNSVII 717
717 MTSNLG---AEHLLAGLTKGVMEVARDCMVEVRKHPRPELNLRLDEIVVDPDLSDHQ 772
718 LTNLSGSGSQAQVLA-----VRATFKPEFINLLDVLIFEGNLNPE 760
773 LRKVARLQKDVAVRLAERGVALLAVTDAAALDYILASYPVYGARPIRMEKVVTELS 832
761 LVRIVDIQLAQGLKRLAQRRLQVSLPAKRWLAQRGFDVYGARFLRLVQQAIGDLA 820
833 KMVREIDENSTVIID 849
|::: ::::|

Db 821 KMLLAGQVHDGDTVPWN 837

RESULT 9

ID CLPB CORGL STANDARD; PRT; 852 AA.
AC P53532;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ClpB protein.
GN CLPB OR CGL2780.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Jaeger W.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RL "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NECESSARY FOR SURVIVAL OF C. GLUTAMICUM AT HIGH TEMPERATURES.
CC -!- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR ATP-DEPENDENT PROTEASE.
CC -!- INDUCTION: By heat shock.
CC -!- SIMILARITY: BELONGS TO THE CLPBA/CLPB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U43536; AAB49540.1; --
DR EMBL; AP005282; BAC00174.1; --
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR001270; Chaprinin_clpA/B.
DR InterPro; IPR004176; Clp_N.
DR Pfam; PF00004; AAA; 2.
DR Pfam; PF02861; Clp_N; 2.
DR PRINTS; PRO0300; CLPPTROTEASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00870; CLPAB 1; 1.
DR PROSITE; PS00871; CLPAB 2; 1.
KW Chaperone; ATP-binding; Repeat; Heat shock; Complete proteome.
FT DOMAIN 162 411 I.
FT DOMAIN 534 725 II.
FT NP_BIND 207 214 ATP (POTENTIAL).
FT NP_BIND 608 615 ATP (POTENTIAL).
SQ SEQUENCE 852 AA; 93231 MW; A00B41BC3DEB30D8 CRC64;

Query Match 47.1%; Score 2145; DB 1; Length 852;
Best Local Similarity 51.4%; Pred. No. 1.1e-86;
Matches 454; Conservative 137; Mismatches 243; Indels 50; Gaps 11;

QY 2 NPEKTHKTNETIATAHELVNAGHAQFTPLHAGALISDPTGTFPPQAISAGGENAQAOS 61
DB 5 NP---TYKQALTAALQAASAGNPIRPAHLMAULTQNDGIAAPLLEAVGVEPATVR 61
QY 62 AERVINOALKKLPSSQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGL 116
DB 62 AE-----AKKLVASYPKASGANMANPNFNRDNLNAFTAAQELAGELGDEYVSTEVLLAG 115

```

117 LLE-DSQIRDLNVEGVATARKVSEKLRGKGGKVESASGDTNFOALKTYGRDLVEQA 175
116 IARGKSDAADDLTKNGATYDAIKEAPPSVRGSG--RVTTQDPGQFOALEKYSTDLTKLA 173
176 --GKLDVPLGRDEIRRVRLISRRTKNNPVLIGBPGVCKTAVVGLAORIVKGVDPNSL 233
174 REGKIDVPLGRDEIRRVVQVLSRRTKNNPVLIGBPGVCKTAVVGLARRIVAGDVPSL 233
234 TDVRLISLDMGALVAGAKYRGFEERLKSVLKEVEDAEAGKVLFDIHLVLCAGKT--EG 292
234 KGKTLISLDSGWSVAGAKYRGFEERLKAVIDEIKGANGEVTFIDELHTIVGASGES 293
293 SMDAANLFPMLARGQLRGICGATTLLEYRYKVEKDAFRRRQQVYVAPSPVDPITSLR 352
294 AMDAGNMKPLLARGELRLVGATTLNEYRYKVEKDAALERRPQQVYVGEPTVEDAIGLR 353
353 GLKEKVEGHGVRIODRALINAQLSARVITGRHLDPKALDVLDAACANRVOLDSPQEE 412
354 GLKEREVHGVRIOQSALVAELSNRIITSRLFPDKAIDLVDEASRLRMEIDSSPOE 413
413 IDNLERKRMQLEIHLALEREKDKASKARLIEVRKELDDRLDKLOPLTMKYRKEKERIDE 472
414 IDELERIVRLEIEEMALSKESDAASKERLEKRLSELADEREKLSLXARWQNEKTAIDD 473
473 IRLKQKRELMFSLQEARRYDLARAADRLYCAIQEVSATAQLEG--TSSEENVMLTEN 531
474 VREMKEELEALSSESIAERDNGNYGRVABRYGRIPLEKQTEADSKVEVVENAMLTSE 533
532 VGPEHIAEVVSWTGPVTRLQNGEKERLIGLADRLHKRVVGQNOAVNAVSEAILRSRAG 591
534 VTPDTIADVVSATGIPAGKMMQGETEKLNNMERNVGNRVGQLEAVTAVSDVRSRAG 593
592 LGRAGQOPTSFLFGPTGVGKTELAKALAEQLFDDENLLVIRIDMSEYMEQHSVRLIGAP 651
594 VADPNRPTGSFLFGPTGVGKTELAKAVAEFLFDDDDRAMIRIDMSEYGEKHSVRLVAGP 653
652 PGVVGHEEGQLTEAVRRPPCYVILFDEVEKHAFAVNTLLQVLDGRLTGQGRVDFR 711
654 PGVVGVDGQQLTEAVRRPPYTVLFDVEVEKAHPDVFLLQVLDGRLTGQGRVDFR 713
712 NSVIIMTSNLGAHLLAGLTGKVTMEVARDCVMRVBRKHFRPELLNRLDEIVDFPLSHD 771
714 NTILLTSLNG-----AGGT-----REQMDAVKAFKPEFVNRLLDDVVDFRLSPE 760
772 QURKVARLQMKDVAVRLAERGVVALAVTDAALDYILAESYDPYVGARPIRRWMEKXVTEL 831
761 QLTISVIDIQIKQLDRLAGRLNRLNRSVDSAKAWLAERGVDPAYGARPLRLIQQAGITL 820
832 SKWVYREEIDENSTYVIDAGADLVYRVESGGLVDASTGKXSDV 875
821 AKELLAGNRDGGVILVDVG-----GQKLDV 848

```

ULT 10

B. TREPA

CLPB TREPA STANDARD; PRG; 878 AA.

083110;

30-MAY-2000 (Rel. 39, Created)

16-OCT-2001 (Rel. 40, Last sequence update)

CLPB protein.

CLPB OR TP0071.

Treponema pallidum.

Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.

NCBI TaxID=160;

[1]

SEQUENCE FROM N.A.

STRAIN: Nichols;

MEDLINE=98332770; PubMed=9665876;

Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,

Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,

Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,

Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,

```

RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
spirochete.";
RL Science 281:375-388(1998).
CC -1- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR
ATP-DEPENDENT PROTEASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001191; AAC65062.1; -.
DR PIR; G71371; G71371.
DR TIGR; TP0071; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_Centr.
DR InterPro; IPR001270; Chaprinin_ClpA/B.
DR InterPro; IPR004176; Clp_N.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF02861; Clp_N; 2.
DR PRINTS; PR00300; CLP_PROTEASIN.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00870; CLPB_1; 1.
DR PROSITE; PS00871; CLPB_2; 1.
DR Chaperone; ATP-binding; Repeat; Complete proteome.
FT DOMAIN 161 409 I.
FT DOMAIN 535 726 II.
FT NP_BIND 206 213 ATP (POTENTIAL).
FT NP_BIND 609 616 ATP (POTENTIAL).
SQ SEQUENCE 878 AA; 79981 MW; 7D9E7419E42A2202 CRC64;

```

Query Match

Best Local Similarity 46.2%; Score 2104; DB 1; Length 878;

Matches 434; Conservative 158; Mismatches 253; Indels 24; Gaps 7;

```

Qy 1 MNEPKTHKNTETIAFAHNAVAGHAQTPPLHAGALISDPTGIPFPAISAGGNAQA 60
Db 1 MNTDRYTVKASEALNDALSLAEAEHNGVVEEHLHALLSKDGIISPLIEKIGAKPDF 60
Qy 61 SAERVNQAALKLPSQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLLE- 119
Db 61 YDE--LLQCLRRKPRVTPGAQTRCAPTLISKACARERLAKNKNQDEYVSCHELLAIS 118
Qy 120 DSQIRDLNVEGVATARKVSEKLRGKGGKVESASGDTNFOALKTYGRDLVEQA--GK 177
Db 119 DSNTARLLHSQITSKTSIAALKDIRG--SKRVTSQDPSTFQCLEKYCRDLTLAREEK 176
Qy 178 LDPVIGRDEIRRVRLISRRTKNNPVLIGBPGVCKTAVVGLAORIVKGVDPNSLTDVR 237
Db 177 IDEVIGRDEIRRVQVLSRRTKNNPVLIGBPGVCKTAVVGLARRIVSGDVPESLKGKR 236
Qy 238 LISLDMGALVAGAKYRGFEERLKSVLKEVEDAEAGKVLFDIHLVLCAGKTEGMDAA 297
Db 237 LLSLDLQALVAGAKYRGFEERLKSVLKEVEDAEAGKVLFDIHLVLCAGKTEGMDAS 296
Qy 298 NLFKPMARGQLRGICGATTLLEYRYKVEKDAAFERRFQQVYVAFSPVDTISILRLK 357
Db 297 NLKLPALARGELRSIGATTLNEYRYKVEKDAALERRFQQVYVCFQPTVEDTIALRLQ 356
Qy 358 YEGHGVTRQDRALINAQLSARYITGRHLDPKALDVLDAACANRVOLDSPQEEIDNLE 417
Db 357 YEVHGVRIKDEALVAATVLSDRYITNRFPLPKAIDLVDAAASRLKMETESQFVLDQVE 416
Qy 418 RKMQLIEIHLALEREKDKASKARLIEVRKELDDRLDKLOPLTMKYRKEKERIDEIRL 477
Db 417 RKLQNLNIEKASLLKESDPASKERLEKELAGFLERRAAMQVQWQNEKGRIESSRYK 476

```

DR	EMBL; X52390; CAA36623.1; -.
DR	EMBL; X52410; CAA36664.1; -.
DR	PIR; C35905; C35905.
DR	InterPro; IPR003593; AAA ATPase.
DR	InterPro; IPR003959; AAA ATPase centr.
DR	InterPro; IPR001270; Chaperin_c1pA/B.
DR	InterPro; IPR004176; Clp_N.
DR	Ffam; PF00004; AAA; 2.
DR	Ffam; PF02861; Clp_N; 2.
DR	PRINTS; PR00300; CLPPTHEASEA.
DR	SMART; SM00382; AAA; 2.
DR	PROSITE; PS00870; CLPAB_1; 1.
DR	PROSITE; PS00871; CLPAB_2; 1.
KW	Chaperone; ATP-binding; Repeat.
FT	DOMAIN 165 413 ;
FT	FT DOMAIN 536 727 ;
FT	NP BIND 210 217 ;
FT	NP BIND 610 617 ;
SQ	SEQUENCE 860 AA; 96271 MW; 34EC8EB5CFC4CF1 CRC64;
Query Match	
Best Local Similarity 44.8%; Score 2039; DB 1; Length 860;	
Matches 423; Conservative 160; Mismatches 255; Indels 12; Gaps 7	
QY	4 EKFTKMTETTAHFLAVNAGHAQFTPLHLACALISDTPGTFPPQAISGAGENAQAASAE 63
DB	8 KFTARFQEALSAAQSGLVAGKHAYIEPLHIFSNLLDEGSSIIATAAGGGIGA--VR 65
QY	64 RVINQALKLPSQPDPDDIPASSSLIKVIRRAQAQKSRGDTHLAVQLIMGLDSQ- 122
DB	66 QAVMQALERLRVNKTPTGDVNIISPESRMNLNLCDKYAQQNGDEIYS-SELFRAVVDAGK 124
QY	123 -IRDLLNEVGVAIVARVXSEVEKLKGEGKVESASGDTNFQALKTVGRDLVEQA--GKLD 179
DB	125 DLEQLLRANGLEAKAAVVAADAVRG--GEAVTDEYAEKGCALKKXTLDTVQRPLDGKD 182
QY	180 PVIGDEIRRVRLSRRTKNPNVLIGEPGVGKTAIVVEGLAQRIYVKGDVPNSLTDVRLI 239
DB	183 PVIGDEIRRAMOILQRRSKNNPVLIGEPGVKTAIVEGLAQRIADRAVPESLKGRLL 242
QY	240 SLDMCALVAGAKYGEFEERLKSVLKEVEDAEKVLFTIDEHLVLGAKTGESMDAANL 299
DB	243 SLDLAALLAGTKYGEFEERLKAVIDEITKADGOIILFDIHTWVGAKSGESLDAGNM 302
QY	300 FKPMIALQRLICGATTLLEVRKYVEKDAFERRFQVVVAEPSVDTTISILRGLKEKYE 359
DB	303 LKPALARGLHCIGATTLDEVQVMKDALEERRFKVIVDESVEDTTIALRNLOERYE 362
QY	360 GHGVRIDRALINAQISARYITGRHLPDKADLVDEACANVTRVOLDSQPEINDLERK 419
DB	363 VHHGINITDPAVAAAQLSHYISGRKLPDKADLVDEAAQAIRMELDSKPEVMOKIDER 422
QY	420 RMQLIELHALEREKDKASKARLIETVRKELDLDLKDQLPTMKRYEKKERIDIERRLKOK 479
DB	423 LIQQLIERMALEKETDAASKRSLDLEAETAQKEYADLEEETLAEKAGNAGAAREIKEQ 482
QY	480 REELMFSIQEAERYIDLARAADLYRGAIOEVESAIAQLE-GTSSENVMLTENVGFPHIA 538
DB	483 LDCLRVELAEAKRGDPARASEIQYGLIPAKEKQLLENEQOTEQRPRLMRNKVTAEEITA 542
QY	539 EVRSRWGTGIPVTRLGONEKERLIGLADRHLKRVVQGQNAVNAVEAILRESAGLGRAQOP 598
DB	543 EIVRSRWGTGIPVAKMMBGEKERLHLLETVLNERVVGOKTAVEAVANAIRNRNLAGLSDPKRP 602
QY	599 TGSFLFGPTGVGKTEALAKABQLDFDENLLVRIDMSYMEQHVSRLIAGPPGVVGHIE 658
DB	603 IGSFLFGPTGVGKTELCTLAQLFDSEENMWVRIDMSFEKHSVARLIAGPPGVVGD 662
QY	659 EGGOITAVRRRPVCVILFOVEBKAHVAVNTLLQVLDDGRITDGGRTVDPRNSVIINT 718
DB	663 QGGYLTAIVARKPYSVVILFOVEBKASHDVNTLLQVLDDGRITDGGRTVDPRHTVIINT 722
QY	719 SNLGAETHLAGLTGKVTMEVARDCWREVRKCHPERPELLNRLEDIVVDPLSHOOLRKVAR 778

723 SNLGS-D-MIOLAEKSYEMKSAWVEIWAHFRPEFINRIDEAIVFHLAKTMYRIAQ 780
779 LQMKDVAFLAERGVALAVTDAALDYILAESVDVPGARPIRRWMEKKVTVLSQWVRE 838
781 IQELRLRQLTRELLSVEEDAINQLVGLDPLFGARPLKRAIQVNIENPLAQALLAG 840

839 EIDENSTVYI 848

841 QVLPQSTITI 850

UNT 12

4 YEAST

H104 YEAST STANDARD; PRT; 908 AA.

P31539;

01-JUL-1993 (Rel. 26, Created)

01-FEB-1996 (Rel. 33, Last sequence update)

15-SEP-2003 (Rel. 42, Last annotation update)

Heat shock protein 104.

HSP104 OR YL026W OR L0948.

Saccharomyces cerevisiae (Baker's yeast).

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

NCBI_TaxID=4932;

[1]

SEQUENCE FROM N.A.

STRAIN=X2180 / ATCC 26109;

MEDLINE=91375541; PubMed=1896074;

Parsell D.A., Sanchez Y., Stitzel J.D., Lindquist S.;

"Hsp104 is a highly conserved protein with two essential nucleotide-

binding sites.";

Nature 353:270-273 (1991).

[2]

SEQUENCE FROM N.A.

STRAIN=S288c / AB972;

MEDLINE=97313267; PubMed=9169871;

Johnston M., Hillier L., Riles L., Albertman K., Andre B., Anseorge W.,

Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,

Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,

Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,

Louis E.J., Messenguy F., Mewes H.-W., Mioaga T., Moestl D.,

Mueller-Auer S., Neutwich U., Obermaier B., Piravandi E., Pohl T.M.,

Portetalle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,

Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,

Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhaesselt P.,

Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,

Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;

"The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";

Nature 387:87-90 (1997).

[3]

SEQUENCE OF 749-908 FROM N.A.

STRAIN=S288c;

MEDLINE=97197984; PubMed=9046100;

Purnelle B., Goffeau A.;

"The sequence of 32kb on the left arm of yeast chromosome XII reveals

six known genes, a new member of the seripauperins family and a new

ABS transporter homologous to the human multidrug resistance

protein.";

Yeast 13:183-188 (1997).

!- FUNCTION: VITAL FOR TOLERANCE TO HEAT, ETHANOL AND OTHER STRESSES.

!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.

!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announcements> or send an email to license@isb.ch).

EMBL; M67479; AAA50477.1; -

DR EMBL; Z73131; CAA97475.1; -
DR EMBL; Z73130; CAA97474.1; -
DR EMBL; X97560; CAA66164.1; -
DR PIR; S61476; S61476.
DR SWISS-2DPAGE; F31539; YEAST.
DR SGD; S0003949; HSP104.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0003767; F:co-chaperone activity; IDA.
DR GO; GO:0003773; F:heat shock protein activity; IDA.
DR GO; GO:0006457; P:protein folding; IDA.
DR GO; GO:0006950; P:response to stress; IDA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR001270; Chaplin_clpA/B.
DR InterPro; IPR004176; Clp_N.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF02861; Clp_N; 2.
DR PRINTS; PR00300; CLIPROTEASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00870; CLPAB_1; 1.
DR PROSITE; PS00871; CLPAB_2; 1.
KW Chaperone; Heat shock; ATP-binding; Repeat.
FT DOMAIN 167 411
FT NP BIND 541 731
FT NP BIND 212 219 ATP (POTENTIAL).
FT NP BIND 614 621 ATP (POTENTIAL).
SQ SEQUENCE 908 AA; 102034 MW; 4AD0E7E3AF98E318 CRC64;

Query Match 41 5%; Score 1890; DB 1; Length 908;
Best Local Similarity 43.2%; Pred. No. 1.6e-75;
Matches 405; Conservative 181; Mismatches 288; Indels 64; Gaps 16;

QY 1 MNPE-KFTHTKNTIATATHELVAVNAGHAQFTPLHLAGALISDP-TGIFPQAISAGGENA 58
DB 1 MNDQTFERALTILTILTAQKLASDHQHPQLQPIHILAAFIETPDGSPYLV-----QNL 54

QY 59 AGSA-----ERVINQALKULPSGPPDDIPASSSLIKVIRRAQAAQKSGDTHLAVD 111
DB 55 IEKGRYDYDLFKVNVNRNIVRIPQQPAPAEITPVALGKVLQDAAKIQKQKQSFIAQD 114

QY 112 QLMGLDLSQIRLDLNEVGAVATARKVEKLRGKGVKESASGDTN--FOALKTYGR 169
DB 115 HILFALFNDSQIFKEAQNVDIEAKQALSELRG--NTRIDSRGADINTPLEYLSKVAI 172

QY 170 DLVQQA--GKLDVIGRDEIRRVRIILSRRTKNNPVLITGEPGVGTAVVEGLAQIRVKG 227
DB 173 DMTEQARQKGLDPVIGRBEIRSTIRVLARIRKSNPNFLIGBPGIKTAIEGVAQRIIDD 232

QY 228 DVPNSLTDVRLISLDMGALVACAKYRGFEERLKSVLKEVEDAEKVLFDIEHLVLA 287
DB 233 DVPTILQGAQLFSLDLAALTAGAKYKGDPEERFKGLKEIBESKTLIVLFDIEHLMGN 292

QY 288 GKTEGSDMAANLFPKMLARGQRCIGATTLEIRYKVEKDAAFERRFOOVVVAEPSVPT 347
DB 293 GKD----DAANILKALSEGQLKVIGATTNRYRIVKDGAFERRFOKIEVAEPSVRQT 348

QY 348 ISILRLKKEKYGHGVRIQDRALNAAQLSARYITGRHLPDKAIDLVDKCANVVRVOLD 407
DB 349 VAILRLQPKYIEIHGVRIILDSALVTAQAQKLYPRRLPDSALDLVDISCAGVAVARD 408

QY 408 SOPERDNLERKEMQLETELHALEK--DKASKARLVEVRKELDDLDKQLPLTMKVRK 465
DB 409 SKEPELDSKERQLQIQIQLALEDEDASTTKDRKLAKQKASLQEELEPLRQRVNE 468

QY 466 EKERIDEIRLRKQKEELMFSLQEAERYDYLRARADLRVGAIQVEASAQLEGTSSEE- 524
DB 469 EKHGHEELTQAKKLDLENKALDAERYDYATAADLRVFAIPDKIKLEKLEQVAAEE 528

QY 525 -----NVMLTENVGPEHIAEVVVRWTGIPVTRLGQNEKERLIGLADRLHKRVVGQNAV 579
DB 529 RRAGANSMTQNVVDTSITSETAARLTGIPVKKLSENEKLIHNERDLSSSEVVGQMDA 588

QY 580 AVSEAILSRAGLGRAQQTGSLFGTGTGKTELAKALAEQLFDDENLIVRDMSEYM 639

RESULT 14
CLPB HELPY

RESULT 14
CLPB HELPY

CLPB HELPY STANDARD; PRT; 856 AA.
 P71404;
 01-NOV-1997 (Rel. 35, Created)
 01-NOV-1997 (Rel. 35, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 ClPB protein.
 CLPB OR HP0264.
 Helicobacter pylori (Campylobacter pylori).
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 Helicobacteraceae; Helicobacter.
 NCBI_TaxID=210;

[1]
 SEQUENCE FROM N.A.
 STRAIN=NCCTC 11637;
 Allan E.;
 Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

[2]
 SEQUENCE FROM N.A.
 STRAIN=26695 / ATCC 700392;
 MEDLINE=97394467; PubMed=9252185;
 Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
 Hayes W.S., Borodovsky M., Kar P.D., Smith H.O., Fraser C.M.,
 Venter J.C.;
 "The complete genome sequence of the gastric pathogen Helicobacter
 pylori.";

Nature 388:538-547(1997).

-!- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR

ATP-DEPENDENT PROTEASE (BY SIMILARITY).

-!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch>)
 or send an email to license@isb-sib.ch.

EMBL; Y08238; CAA69406.1; -.
 EMBL; AE000545; AAD07330.1; -.
 PIR; H64552; H64552.

TIGR; HP0264; -.

InterPro; IPR003593; AAA_Atpase.

InterPro; IPR003599; AAA_Atpase_cent.

InterPro; IPR001270; Chaprinin_clpa/B.

InterPro; IPR004176; Clp_N.

Pfam; PF00004; AAA; 2.

Pfam; PF02861; Clp_N; 2.

PRINTS; PR00300; CLP_PTEASEA.

SMART; SM00382; AAA; 2.

PROSITE; PS00870; CLPB_1; 1.

PROSITE; PS00871; CLPB_2; 1.

Chaperone; ATP-binding; Repeat; Complete proteome.

DOMAIN 159 406 I.

DOMAIN 531 722 II.

NP_BIND 204 211 ATP (POTENTIAL).

NP_BIND 605 612 ATP (POTENTIAL).

CONFLICT 14 14 T -> A (IN REF. 1).

CONFLICT 33 33 M -> L (IN REF. 1).

CONFLICT 59 59 Q -> E (IN REF. 1).

CONFLICT 62 62 R -> K (IN REF. 1).

CONFLICT 79 79 S -> N (IN REF. 1).

CONFLICT 103 103 R -> T (IN REF. 1).

CONFLICT 119 119 G -> S (IN REF. 1).

CONFLICT 131 131 A -> T (IN REF. 1).

CONFLICT 145 145 R -> A (IN REF. 1).

CONFLICT 149 149 D -> G (IN REF. 1).

FT	CONFLICT	222	222	M -> V (IN REF. 1).
FT	CONFLICT	435	435	A -> H (IN REF. 1).
FT	CONFLICT	465	465	V -> A (IN REF. 1).
FT	CONFLICT	644	644	MS -> IT (IN REF. 1).
FT	CONFLICT	649	649	A -> P (IN REF. 1).
FT	CONFLICT	734	734	E -> D (IN REF. 1).
FT	CONFLICT	765	765	D -> G (IN REF. 1).
FT	CONFLICT	836	836	V -> I (IN REF. 1).
FT	CONFLICT	840	840	D -> G (IN REF. 1).
SQ	SEQUENCE	856	856	AA; 96683 MW; E3902C7D989496AD CRC64;

Query Match 41.2%; Score 1876.5; DB 1; Length 856;
 Best Local Similarity 45.0%; Pred. No. 5.8e-75;
 Matches 390; Conservative 191; Mismatches 262; Indels 23; Gaps 11;

Qy	4	EKTHKNETIATHELAHVNAHQAQTPLHAGALISDFTGIPPOAISAGGENAAQSAE	63
Db	5	EKMTDQLHETLSALALHKKNAEVTMHLFAMLNSQGILIQALQMPVD--IQALR	62
Qy	64	RVINOALKLPSPSP--PPDIPASSSLIKVIRRAQAQKSRGDTHLAVDOLI--MGLLE	119
Db	63	LSVQSELNKFAKVSQISKQNIQLNQALISLENAQGLMAKRGDSFIATDVYLLANGLFE	122
Qy	120	DSQIRDLNNEVGAVATARKSEVEKLRGKKGKVESASGDTNFQALKTYGRDLVEQA--GK	177
Db	123	-SVLKPYLD----AKELQKTLBSLR--KGRTTQDKNDSDNLSLEKFGDILTKALENK	174
Qy	178	LDPVIGRDEIRVVRILSRPTKNNPVILGEPGVGKTAVVEGLAQRIKVGDPVNSLTDVR	237
Db	175	LBPVIGRDEIRVVRILSRPTKNNPVILGEPGVGKTAVVEGLAQRIKVGDPVNSLTDVR	234
Qy	238	LISLDGALVAGAKYRGEFEERLKSVLKVEVDAEGKVLFIIDEIHLVLGAGTSGMDAA	297
Db	235	VIALDLSLVAGAKYRGEFEERLKSVLKVEVDAEGKVLFIIDEIHLVLGAGTSGMDAA	293
Qy	298	NLPKPMARQRCIGATTELYRKYVEKDAAPERFQVYVAESPPTIISILRLGKKEK	357
Db	294	NILKPALARGELHTIGATTELYRKYVEKDAAPERFQVYVAESPPTIISILRLGKKEK	353
Qy	358	YEGHGVRIQDRALINAQSLARYITGRHLPKADLVDDEACANVRVQLDSQPEIDNLE	417
Db	354	LETHNITINDSALIASKLSRYITDFLPKADLIDDEGAAQLKMQMSEPAKLISYVK	413
Qy	418	RKMQLIEIHLALEREKOKASKARLFEVRKELDDLRDKLQPLTKMYRKEKERIDIRRLK	477
Db	414	RSTQRLMEKQALEMEKESNAKQMEIILKELSDLKKEKIQLEAQFENEKEVKEISRLK	473
Qy	478	QKREELMPSLOEAREYDYLARAADLYGAIQEVESAIAQLE--GTSSEENVMLTENVGPE	535
Db	474	MEMESLKKEAERFRNGDYQQAAGEIYSKIPENKKEBELQKWEAMQONGALLQNALTE	533
Qy	536	-HTAEVSVRWGIPVTRLGONKEKELIGLADRLHKVVGONAVNAVSEAILRSRAGLGR	594
Db	534	NNIAEIVSQWTHIPVQKMLQSEKNRVINTESELRKRVGQEKATKAIAKAIKRNKAGLS	593
Qy	595	AOQPTGSFLPGTGVGKTELAKALAEQLFDDENLIVRIDMSEYMEQHSVSLIGAPGY	654
Db	594	SNKPIGSFLPGTGVGKTESAKALAEQLFDDENLIVRIDMSEYMEQHSVSLIGAPGY	653
Qy	655	VGHEGGQLTAVRRRPYCVILFDEVEKARHVAVENTLLQVLDGRLTGGQRTVDFRNSV	714
Db	654	VGYEGGQLTAVRRRPYCVILFDEVEKARHVAVENTLLQVLDGRLTGGQRTVDFRNSV	713
Qy	715	IIMTSNLGAHLLAGITGVKVTMEVARDVVRVVRKHPPELNLDELIVVDFPLSHDLR	774
Db	714	LIITSNVASGALLEENLSEAKQKA---IKESLRQFFKPEFLNRLDEIISFNALDSHAVI	770
Qy	775	KVARLQMDVAVRLAERGVAVLTAADLYILAESYDYPYGVGPIRRRWMEKKVVTLSKM	834
Db	771	NIVGILFENIQQKALERGINITLDBEAKELIAEAGDFRYPGARPLKRALYEWEDKLAEL	830
Qy	835	VUREEDENSTVVIDAGAGDIIVRYE	860

Search completed: February 13, 2004, 01:33:14
Job time : 44 BECS

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: February 12, 2004, 22:31:20 ; Search time 73 Seconds
(without alignments)
3220.355 Million cell updates/sec

le: US-09-812-350-17
fect score: 4550
uence: 1 MNPEKFKHKTETATAHEL.....VKQWRIBEDDDNEEMIED 911

ring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

rchd: 830525 seqs, 258052604 residues

al number of hits satisfying chosen parameters: 830525

imum DB seq length: 0
imum DB seq length: 2000000000
t-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Alt	Score	Query Match	Length	DB ID	Description
1	4003	88.0	911	10	Q39889 glycine max
2	3896	85.6	909	10	Q92T13 nicotiana t
3	3883.5	85.4	912	10	Q98822 zea mays t
4	3840.5	84.4	912	10	Q8W2B5 oryza sativ
5	3825	84.1	918	10	Q9XE11 oryza sativ
6	3822.5	84.0	913	10	Q9SPH4 oryza sativ
7	3807.5	83.7	912	10	Q8L6I4 oryza sativ
8	3790	83.3	918	10	Q92T12 triticum ae
9	2502.5	55.0	582	10	Q9SY59 zea mays (m
10	2431	53.4	668	10	Q23323 arabidopsis
11	2299.5	50.5	871	16	Q8DJ40 synechococc
12	2298	50.5	867	5	O00828 leishmania
13	2294	50.4	931	16	Q8YJ91 bruceella me
14	2293	50.4	874	16	Q9AEM5 bruceella su
15	2289	50.3	867	5	Q25317 leishmania
16	2258.5	49.6	865	16	Q97KG0 clostridium

17	2242.5	49.3	866	16	Q929G7
18	2238	49.2	866	16	Q8XKG8
19	2236.5	49.2	866	16	Q8Y570
20	2234.5	49.1	905	3	O94641
21	2225	48.9	868	16	Q98G96
22	2224	48.9	859	16	Q9A9T4
23	2218.5	48.8	835	16	Q8YM56
24	2217.5	48.7	883	2	O87444
25	2209	48.5	868	16	Q92MK7
26	2204	48.4	857	16	Q8XFM5
27	2199	48.3	887	16	Q8U8B5
28	2197	48.3	855	16	Q8CKC0
29	2185	48.0	889	16	O8G4X4
30	2184	48.0	852	16	Q8FM94
31	2182.5	48.0	898	16	P74459
32	2180.5	47.9	871	16	Q8PHQ4
33	2174.5	47.8	861	16	Q8P6A0
34	2168.5	47.7	858	16	Q92EA9
35	2166	47.6	968	10	O9LF37
36	2165.5	47.6	865	16	O8CJV9
37	2163.5	47.5	857	16	Q92JK8
38	2161	47.5	864	16	Q8D181
39	2160	47.5	857	16	Q8ZBV9
40	2158.5	47.4	861	16	Q9PGC1
41	2158.5	47.4	877	2	Q9ZAX2
42	2151	47.3	857	2	Q9KHC1
43	2150.5	47.3	848	16	Q9CB26
44	2144.5	47.1	864	16	Q8RH08
45	2141.5	47.1	869	16	Q99VB5

ALIGNMENTS

RESULT 1

Q39889 PRELIMINARY; PRT; 911 AA.
ID Q39889
AC Q39889;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Heat shock protein.
GN SB100.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroside 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95170290; PubMed=7866031;
RA Lee Y.R., Nagao R.T., Key J.L.;
RT "A soybean 101-KD heat shock protein complements a yeast HSP104
deletion mutant in acquiring thermotolerance.";
RL Plant Cell 6:1889-1897(1994).
DR EMBL; L35272; AAA66338.1; -
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR001270; Chaprinin_cipa/B.
DR Pfam; PF00004; AAA; 2.
DR Pfam; PF02861; Clp_N; 2.
DR PRINTS; PR00300; CLP_PTEASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00870; CLPAB_1; 1.
DR PROSITE; PS00871; CLPAB_2; 1.
KW ATP-binding.
SQ SEQUENCE 911 AA; 101328 MW; 6A2E1EE309067884 CRC64;

Query Match 88.08; Score 4003; DB 10; Length 911;
Best Local Similarity 87.08; Pred. No. 1.3e-187;
Matches 794; Conservative 72; Mismatches 39; Indels 8; Gaps 6;

1 MNPEKFTHTKNTTATATAHELAVNAGHAQFTPLHLAALISDPTGIFPQAISSA-GGENAA 59
 1 MNPEKFTHTKNTTATATAHELAVNAGHAQFTPLHLAALISDPTGIFPQAISSA-GGENAA 60
 60 QSAERVINQALKKLPSPQPPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQILGLLE 119
 61 RAVERVNLQALKKLPSPQPPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQILGLLE 120
 120 DSQIRDLNVEGVATARKVSEVEKLRGKGGKVESASGDTNFQALTKYGRDLVQAGKLD 179
 121 DSQIGDLLKEAGVAVAKVSEVVDKLRGKGGKVESASGDTNFQALTKYGRDLVQAGKLD 180
 180 PVIGRDEIRRVRLSRRTKNNPVILGEGVGTAVVEGLAQRIKVGDPVNSLTDVRLI 239
 181 PVIGRDEIRRVRLSRRTKNNPVILGEGVGTAVVEGLAQRIKVGDPVNSLTDVRLI 240
 240 SLDGALVAGAKYGEFEERLKSILKEVEDEAGKVLFIIDEIHLVLAGKTEGSDMAANL 299
 241 ALDMGALVAGAKYGEFEERLKSILKEVEDEAGKVLFIIDEIHLVLAGKTEGSDMAANL 300
 300 FKMLARGQLRCIGATTLEIRYKVEKDAAFERRFOQVVAEPPVDTISILRLKERYE 359
 301 FKMLARGQLRCIGATTLEIRYKVEKDAAFERRFOQVVAEPPVDTISILRLKERYE 360
 360 GHGVRIQDRALINAAQLSARYITGRLPKDAIDLVDACANVRVOLDSPQEIIDLNRK 419
 361 GHGVRIQDRALINAAQLSARYITGRLPKDAIDLVDACANVRVOLDSPQEIIDLNRK 420
 420 RMQLEIHALEREKDKASKARLIEVRKELDLRLKQPLMKYRKEKERIDEIRLKK 479
 421 RMQLEIHALEREKDKASKARLIEVRKELDLRLKQPLMKYRKEKERIDEIRLKK 480
 480 REELMFSLOEAEERYDLARAADLYGAIQEVESAIQAQLEGTSSEENVMLTENVPETIAE 539
 481 REELMFSLOEAEERYDLARAADLYGAIQEVESAIQAQLEGTSSEENVMLTENVPETIAE 539
 540 VVSRWTGIPVTRLGQNEKERLIGLADRLHKRVVGVQNAVNSEALIRSRAGLGRAQPT 599
 540 VVSRWTGIPVTRLGQNEKERLIGLADRLHKRVVGVQNAVNSEALIRSRAGLGRAQPT 599
 600 GSPFLGPTGVGTALAKALAEQFDENQLVRIDMSYMEQHSVRLIGAPGVVGHHEE 659
 600 GSPFLGPTGVGTALAKALAEQFDENQLVRIDMSYMEQHSVRLIGAPGVVGHHEE 659
 660 GQGLTEAVRRPVCVILFDEVEKAVHVAFTLLQVLDGRLTDGQRTVDFNSVIMTS 719
 660 GQGLTEAVRRPVCVILFDEVEKAVHVAFTLLQVLDGRLTDGQRTVDFNSVIMTS 719
 720 NLGAELHLAGLTKVMEVARCVMEVRKRPPELLNRLDEIVVDFPLSHDQLEKVARL 779
 720 NLGAELHLAGLTKVMEVARCVMEVRKRPPELLNRLDEIVVDFPLSHDQLEKVARL 779
 780 QMKDVAVRLAERGVAVTDAALDYLAESYDPVYGARPIRMWKEKKVVTLSKRVVREE 839
 780 QMKDVAVRLAERGVAVTDAALDYLAESYDPVYGARPIRMWKEKKVVTLSKRVVREE 839
 840 IDENSTVYIDAG--AGDLAVRYE--SGGLVDASTGKSDVILHIANG--PKSDAAQAVKK 894
 840 IDENSTVYIDAG--AGDLAVRYE--SGGLVDASTGKSDVILHIANG--PKSDAAQAVKK 894
 895 MRIEIEDDDNEE 907
 899 MKIEIEDDDNEE 911

SULT 2
 Q92T13
 Q92T13
 PRELIMINARY; PRT; 909 AA.
 01-MAY-1999 (Tremblrel. 10, Created)
 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 101 kDa heat shock protein.

GN HSP101.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9903076; PubMed=9784498;
 RA Wells D.R., Tanguay R.L., Le H., Gallie D.R.;
 RT "HSP101 functions as a specific translational regulatory protein whose
 activity is regulated by nutrient status.";
 RL Genes Dev. 12:3236-3251(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Gallie D.R., Tanguay R.L., Wells D.R.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF083343; AAC83688.2; -
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR001270; Chaprinin_c1pA/B.
 DR InterPro; IPR004176; Clp_N.
 DR Pfam; PF00004; AAA; 2.
 DR Pfam; PF02861; Clp_N; 2.
 DR PRINTS; PR00300; CLPPTASEA.
 DR SMART; SM00382; AAA; 2.
 DR PROSITE; PS00870; CLPAB_1; 1.
 DR PROSITE; PS00871; CLPAB_2; 1.
 KW ATP-binding; Heat shock.
 SQ SEQUENCE 909 AA; 101110 MW; 86641AD7FA47678D CRC64;

Query Match 85.6%; Score 3896; DB 10; Length 909;
 Best Local Similarity 85.0%; Pred. No. 2.3e-182;
 Matches 774; Conservative 75; Mismatches 56; Indels 6; Gaps 5;
 QY 1 MNPEKFTHTKNTTATATAHELAVNAGHAQFTPLHLAALISDPTGIFPQAISSA-GGENAA 59
 DB 1 MNPEKFTHTKNTTATATAHELAVNAGHAQFTPLHLAALISDPTGIFPQAISSA-GGENAA 60
 QY 60 QSAERVINQALKKLPSPQPPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQILGLLE 119
 DB 61 NSERVNLQALKKLPSPQPPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQILGLLE 120
 QY 120 DSQIRDLNVEGVATARKVSEVEKLRGKGGKVESASGDTNFQALTKYGRDLVQAGKLD 179
 DB 121 DSQIGDLLKEAGVAVAKVSEVVDKLRGKGGKVESASGDTNFQALTKYGRDLVQAGKLD 180
 QY 180 PVIGRDEIRRVRLSRRTKNNPVILGEGVGTAVVEGLAQRIKVGDPVNSLTDVRLI 239
 DB 181 PVIGRDEIRRVRLSRRTKNNPVILGEGVGTAVVEGLAQRIKVGDPVNSLTDVRLI 240
 QY 240 SLDGALVAGAKYGEFEERLKSILKEVEDEAGKVLFIIDEIHLVLAGKTEGSDMAANL 299
 DB 241 ALDMGALVAGAKYGEFEERLKSILKEVEDEAGKVLFIIDEIHLVLAGKTEGSDMAANL 300
 QY 300 FKMLARGQLRCIGATTLEIRYKVEKDAAFERRFOQVVAEPPVDTISILRLKERYE 359
 DB 301 FKMLARGQLRCIGATTLEIRYKVEKDAAFERRFOQVVAEPPVDTISILRLKERYE 360
 QY 360 GHGVRIQDRALINAAQLSARYITGRLPKDAIDLVDACANVRVOLDSPQEIIDLNRK 419
 DB 361 GHGVRIQDRALINAAQLSARYITGRLPKDAIDLVDACANVRVOLDSPQEIIDLNRK 420
 QY 420 RMQLEIHALEREKDKASKARLIEVRKELDLRLKQPLMKYRKEKERIDEIRLKK 479
 DB 421 RMQLEIHALEREKDKASKARLIEVRKELDLRLKQPLMKYRKEKERIDEIRLKK 480
 QY 480 REELMFSLOEAEERYDLARAADLYGAIQEVESAIQAQLEGTSSEENVMLTENVPETIAE 539
 DB 481 REELMFSLOEAEERYDLARAADLYGAIQEVESAIQAQLEGTSSEENVMLTENVPETIAE 539
 QY 540 VVSRWTGIPVTRLGQNEKERLIGLADRLHKRVVGVQNAVNSEALIRSRAGLGRAQPT 599
 DB 540 VVSRWTGIPVTRLGQNEKERLIGLADRLHKRVVGVQNAVNSEALIRSRAGLGRAQPT 599

540 VVSRWTGIPVSRIGQNEKEKLGDLRHLQRVVGQDHAVRAVAEAVLSRAGLRPOQPT 599
600 GSFLGPTGKGTAKALAEQLEDDENLLVRIDMSEYMEQHSVSRILGAPPGVVGHEE 659
600 GSFLGPTGKGTAKALAEQLEDDENLLVRIDMSEYMEQHSVSRILGAPPGVVGHEE 659
660 GQGLTEAVRRRYPYCVILFDEVEKAVHVAVENTLLQVLDGRLTDGQRTVDFRNSVIIMTS 719
660 GQGLTEAVRRRYPYCVILFDEVEKAVHVAVENTLLQVLDGRLTDGQRTVDFRNSVIIMTS 719
720 NGAELHLAGLTKVMEYARDVCMVREKHPPELLNRDLDEIVVDFPLSHDQLRKVARL 779
720 NLGAELHLAGLTKVMEYARDVCMVREKHPPELLNRDLDEIVVDFPLSHDQLRKVARL 779
780 QMKDVALRAERGAVALAVDAALDVLAEVDPVYGARPIRRMMEKKVVTLSKQVVRRE 839
780 QMKDVALRAERGAVALAVDAALDVLAEVDPVYGARPIRRMMEKKVVTLSKQVVRRE 839
840 IDENSTVVIDAGAG--DLVYRVE--SGGLVDASTGKSDVLHIANGPKRSAAQAVKQK 896
840 IDENSTVVIDAGAG--DLVYRVE--SGGLVDASTGKSDVLHIANGPKRSAAQAVKQK 896
897 IEIEDDDNEE 907
899 IEIEDDDMED 909

ULF 3
8822
Q98822 PRELIMINARY; PRT; 912 AA.
Q98822; 01-MAY-2000 (TremBLrel. 13, Created)
01-MAY-2000 (TremBLrel. 13, Last sequence update)
01-JUN-2002 (TremBLrel. 21, Last annotation update)
Heat shock protein HSP101.
HSP101.
Zea mays (Maize).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
PACCAD clade; Panicoidae; Andropogoneae; Zea.
NCBI_TaxID=4577;
(1)
SEQUENCE FROM N.A.
Young T.E., Gallie D.R.;
"HSP101 Diversity in Monocot Species";
Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
(2)
SEQUENCE FROM N.A.
MEDLINE=99234099; PubMed=10216257;
Nieto-Sotelo J., Kannan K.B., Martinez L.M., Segal C.;
"Characterization of a maize heat shock protein 101 gene, HSP101,
encoding a ClpB/Hsp100 protein homologue";
Gene 230:187-195 (1999).
EMBL; AF133840; AAD33606.1; -;
EMBL; AF077337; AAD25223.1; -;
InterPro; IPR003593; AAA_ATPase.
InterPro; IPR003959; AAA_ATPase_cent.
InterPro; IPR001270; Chaprinin_c1pA/B.
InterPro; IPR004176; Clp_N.
Pfam; PF00004; AAA; 2.
Pfam; PF02861; Clp_N; 2.
PRINTS; PR00300; CLP_N; 2.
SMART; SM00382; AAA; 2.
PROSITE; PS00870; CLPAB_1; 1.
PROSITE; PS00871; CLPAB_2; 1.
ATP-binding; Heat shock.
SEQUENCE 912 AA; 101131 MW; 13FIEAA4BEA610E2 CRC64;
very Match 85.4%; Score 3893.5; DB 10; Length 912;
Best Local Similarity 84.6%; Pred. No. 9.4e-182;
Matches 772; Conservative 76; Mismatches 58; Indels 7; Gaps 6;
1 MNPEKTHKNTIATAHAVAGHQAQTPLHLAALISDPTGIPFQAISGA-GENNA 59

1 MNPEKTHKNTIATAHAVAGHQAQTPLHLAALISDPTGIPFQAISGA-GENNA 60
60 -QSAERYINQALKKLPSPSPDDIPASSSLIKVIRAAQAKSRGTHLAVDQIMGLL 118
61 GDSFERVNNLSKKLPSPSPDDIPASSSLIKVIRAAQAKSRGTHLAVDQIMGLL 120
119 EDSQIRDLNVEGVATARKVEKLEKGGKGVESAGDTNFQALKTYGRDLVEQAGKL 178
121 EDSQISDLCEAGVSAARVAELEKLGEGRRVVSASGDTNFQALKTYGRDLVEQAGKL 180
179 DPVIGRDEIRRVVRLSRRTKNNPVLIGEPGVGKTAVVEGLAQIRVKGDPVNSLTDVRL 238
181 DPVIGRDEIRRVVRLSRRTKNNPVLIGEPGVGKTAVVEGLAQIRVKGDPVNSLTDVRL 240
239 ISLDMGALVAGAKYGFEEERLKSILKEVEDABGKVLFIIDEHLVLAGKTEGSDAAN 298
241 ISLDMGALVAGAKYGFEEERLKSILKEVEDABGKVLFIIDEHLVLAGKTEGSDAAN 300
299 LFKPMLARGQLRCIGATTLEEVRYKVEKDAAPERRFQVVAEPSPDTISILRGKEKY 358
301 LFKPMLARGQLRCIGATTLEEVRYKVEKDAAPERRFQVVAEPSPDTISILRGKEKY 360
359 EGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVEACANVRVQLDSQPEEDNLER 418
361 EGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVEACANVRVQLDSQPEEDNLER 420
419 KQMLELHLEHLEKDKASKARLIEVRKELDDLDKLOPLTKYRKEKERIDEIRRLKQ 478
421 KRIQLEVLHLEHLEKDKASKARLIEVRKELDDLDKLOPLTKYRKEKERIDEIRRLKQ 480
479 KREELMFSIQEAERYRLARAAADLRVGAIOEVESAIQLEGTSSEENVMLTENVPHEITA 538
481 RREELQFTLQEAERRMDLARVADLYKGLQELDAISKLE-SETGENMLTETVGEPIA 539
539 EVSVRWGTGIPVTRLGQNEKERLIGLADRLHKEVVGQNAVNAVSEAILRSRAGLRQAP 598
540 EVSVRWGTGIPVTRLGQNEKERLIGLADRLHKEVVGQNAVNAVSEAILRSRAGLRQAP 599
599 TGSFLGPTGKGTAKALAEQLEDDENLLVRIDMSEYMEQHSVSRILGAPPGVVGHE 658
600 TGSFLGPTGKGTAKALAEQLEDDENLLVRIDMSEYMEQHSVSRILGAPPGVVGHE 659
659 EGGQTEAVRRRYPYCVILFDEVEKAVHVAVENTLLQVLDGRLTDGQRTVDFRNSVIIMT 718
660 EGGQTEAVRRRYPYCVILFDEVEKAVHVAVENTLLQVLDGRLTDGQRTVDFRNSVIIMT 719
719 SNLGAELHLAGLTKVMEYARDVCMVREKHPPELLNRDLDEIVVDFPLSHDQLRKVAR 778
720 SNLGAELHLAGLTKVMEYARDVCMVREKHPPELLNRDLDEIVVDFPLSHDQLRKVAR 779
779 LQMKDVALRAERGAVALAVDAALDVLAEVDPVYGARPIRRMMEKKVVTLSKQVVRRE 838
780 LQMKDVALRAERGAVALAVDAALDVLAEVDPVYGARPIRRMMEKKVVTLSKQVVRRE 839
839 EIDENSTVVIDAGAG--DLVYRVE--SGGLVDASTGKSDVLHIANGPKRSAAQAVKQK 895
840 EIDENSTVVIDAGAG--DLVYRVE--SGGLVDASTGKSDVLHIANGPKRSAAQAVKQK 896
896 RI-IEIEDDDNEE 907
900 RIMEEDDGMDEE 912

RESULT 4
Q98285 PRELIMINARY; PRT; 912 AA.
ID Q98285
AC Q98285;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)
DE Heat shock protein.
GN HSP101.

Oryza sativa (Rice).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehretidae; Oryzae; Oryza.

NCBI_TaxID=4530;

[1]

SEQUENCE FROM N.A.

STRAIN=cv. Nipponbare;

Agarwal M., Sahi C., Young T., Agarwal S., Mitra S., Ganesan K.,

Agarwal-Katiyar S., Gallie D.R., Grover A., "

"Isolation and Analysis of HSP101 from Rice."

Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

EMBL: AF332981; AAL57165.1;

Gramene; O8W2B5;

InterPro; IPR003593; AAA_ATPase.

InterPro; IPR003959; AAA_ATPase_cent.

InterPro; IPR001270; Chaprinin_clpA/B.

InterPro; IPR004176; Clp_N.

Pfam; PF00004; AAA; 2.

PRINTS; PR00300; CLPPTASEA.

SMART; SM00382; AAA; 2.

PROSITE; PS00870; CLPAB_1; 1.

PROSITE; PS00871; CLPAB_2; 1.

ATP-binding.

SEQUENCE 912 AA; 100896 MW; 0A3B17065B7CA343 CRC64;

Query Match 84.4%; Score 3840.5; DB 10; Length 912;

Best Local Similarity 83.8%; Pred. No. 1.2e-179;

Matches 765; Conservative 75; Mismatches 66; Indels 7; Gaps 5;

1 MNPKFTHKTNETTATATHELVANAGHAQFTPLHLAGALISDPTGIFPQAISAGGENAA- 59

1 MNPNFTHKTNEALVAAHEIASEAGHAQLTPLHVAALADKGGILFQAISQASGGDAGA 60

60 -QSAERVINOALKLPQSPPDDIPASSLSIKVIRAAQAQSRGTHLAVDQLMGLL 118

61 PDSFVVGALKLPQSPPDDIPASTALIKVIRRAQAQSRGTHLAVDQLMGLL 120

119 EDSQIRDLNVEGVATAREKSEKLEKGEKVESAGDTNFQALTKYGRDLVEQAQGL 178

121 EDSLISDLKEAGVSAARVAELKRGEGKVESAGDTNFQALTKYGRDLVEQAQGL 180

179 DPVIGRDETRVRVRLISRTKNNPVLIGEGVGTAVVEGLAQIRIVKGVDPNSLTDVRL 238

181 DPVIGRDETRVRVRLISRTKNNPVLIGEGVGTAVVEGLAQIRIVKGVDPNSLTDVRL 240

239 ISLWGMALVAGAKYRGEPEERLKSVLKEVEDAEKVLFTDEIHLVLGAKTSGMDAAN 298

241 IALWGMALVAGAKYRGEPEERLKSVLKEVEDAEKVLFTDEIHLVLGAKTSGMDAAN 300

299 LFKPLMARGQLRCIGATTLEBYRYKYVEKDAAFERRFQVYVAEPSVPTTISILRGLKEKY 358

301 LFKPLMARGQLRCIGATTLEBYRYKYVEKDAAFERRFQVYVAEPSVPTTISILRGLKEKY 360

359 EGHGVRIQDRALNAAQLSARYTGRHLPKADLDLVEACANRVQLDSQPEIDMLER 418

361 EGHGVRIQDRALNAAQLSARYTGRHLPKADLDLVEACANRVQLDSQPEIDMLER 420

419 KRMQLIEELHALEREKDKASKARLIEYRKEKDLDDLDKLOPLTKMYRKEKDEIRLQK 478

421 KRIQVEVHHALEKEKASKARLIEYRKEKDLDDLDKLOPLTKMYRKEKDEIRLQK 480

479 KREELMFSQBAERYDILARAADLYRGAIOEVESAIAQLEGTSEENVMLTENVPGEHIA 538

481 KREELQFTQBAERMQLARVADLYKALQEIADVAIAKLE-SETGENMLTETVGPGEQIA 539

539 EVVSRWTGIPVTRIGQNEKERLIGLADLRHVRVGVQGNVAVSEAILRSRAGLGRAQOP 598

540 EVVSRWTGIPVTRIGQNEKERLIGLADLRHVRVGVQGNVAVSEAILRSRAGLGRAQOP 599

599 TGSFLFLGPTGVTGKTELAKALAEQLFDDENLLVRIDMSYMEQHSVRLIGAPPGYVGHE 658

600 TGSFLFLGPTGVTGKTELAKALAEQLFDDENLLVRIDMSYMEQHSVRLIGAPPGYVGHE 659

659 EGGQTEAVRRPYCVIILFDEVEKARHVAVENTLLQVLDGRLTGGRTVDFRNSVIIMT 718

660 EGGQTEAVRRPYCVIILFDEVEKARHVAVENTLLQVLDGRLTGGRTVDFRNSVIIMT 719

719 SNLGAELHLAGTGTMTBVARDCVMREVRKHFRPELLNRLDEIVVFDPLSHDQRLKVAR 778

720 SNLGAELHLAGTGTMTBVARDCVMREVRKHFRPELLNRLDEIVVFDPLSHDQRLKVAR 779

779 LOMKDVAVRLAERGVALATDAALDIYLAESYDVPYVGARPIRRMEKVKVWVLSKVVRE 838

780 LOMKDVAVRLAERGVALATDAALDIYLAESYDVPYVGARPIRRMEKVKVWVLSKVVRE 839

839 EIDENSTVYIDAG--AGDLVVRVES--GGLVDASTGKSDVLIHANGPKR--SDAAQAVKK 894

840 EIDENSTVYIDAG--AGDLVVRVES--GGLVDASTGKSDVLIHANGPKR--SDAAQAVKK 899

895 MREIEIEDDDNEE 907

900 KRMEDEDEGMDDE 912

RESULT 5

Q9XE11

ID Q9XE11

AC Q9XE11

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Heat shock protein 101.

GN HSP101A.

OS Triticum aestivum (Wheat).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;

OC Triticeae; Triticum.

OX NCBI_TaxID=4565;

[1]

SEQUENCE FROM N.A.

RC STRAIN=cv. Mustang;

RX MEDLINE=21240210; PubMed=11342108;

RA Campbell J.I., Klueva N.Y., Zheng H.G., Nieto-Sotelo J., Ho T.H.D.,

Nguyen H.T.;

RT "Cloning of new members of heat shock protein HSP101 gene family in

RT wheat (Triticum aestivum (L.) Moench) inducible by heat, dehydration,

RT and ABA."

RL Biochim. Biophys. Acta 1517:270-277(2001).

RL EMBL; AF097363; RAD22629.1; -

DR InterPro; IPR003593; AAA_ATPase.

DR InterPro; IPR003959; AAA_ATPase_cent.

DR InterPro; IPR001270; Chaprinin_clpA/B.

DR InterPro; IPR004176; Clp_N.

DR Pfam; PF00004; AAA; 2.

DR PRINTS; PR00300; CLPPTASEA.

DR SMART; SM00382; AAA; 2.

DR PROSITE; PS00870; CLPAB_1; 1.

DR PROSITE; PS00871; CLPAB_2; 1.

KW ATP-binding.

SQ SEQUENCE 918 AA; 101122 MW; 354D3BAC06C8556C CRC64;

Query Match 84.1%; Score 3825; DB 10; Length 918;

Best Local Similarity 82.6%; Pred. No. 6.9e-179;

Matches 760; Conservative 82; Mismatches 66; Indels 12; Gaps 7;

QY 1 MNPKFTHKTNETTATATHELVANAGHAQFTPLHLAGALISDPTGIFPQAISAGGENAA- 59

1 MNPNFTHKTNEALVAAHEIASEAGHAQLTPLHVAALADKGGILFQAISQASGGDAGA 60

60 -QSAERVINOALKLPQSPPDDIPASSLSIKVIRRAQAQSRGTHLAVDQLMGLL 118

61 PDSFVVGALKLPQSPPDDIPASTALIKVIRRAQAQSRGTHLAVDQLMGLL 120

119 EDSQIDLLNEGVATARKVSEVKLRGKE-GKKVESASGDTNFQALTKYGRDLVEQAGK 177
 121 EDAQADCLKEAGVSASRVRAELEKLRGGDNKRKVESASGDTNFQALTKYGRDLVEQAGK 180
 178 LDPVIGRDEIRRVVRLGRRRTKNNPVLVIGPGVGTAVVEGLAQRIKVGDPVNSLTDVR 237
 181 LDPVIGRDEIRRVVRLGRRRTKNNPVLVIGPGVGTAVVEGLAQRIKVGDPVNSLTDVR 240
 238 LISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEGKVILFIDEIHLVLGAGTEGSMDDAA 297
 241 LVALDMGALVAGAKYRGEFEERLKAFLKEVEAEGKVILFIDEIHLVLGAGTEGSMDDAA 300
 298 NLFKPMALRGQLRCIGATTLEBYRKVKVEKDAFERRFQOYVVAEVSVPDTISILGLKEK 357
 301 NLFKPMALRGQLRCIGATTLEBYRKVKVEKDAFERRFQOYVVAEVSVPDTISILGLKEK 360
 358 YEGHGVRIQDRALINAAQLSARYITGRHLDPKADLVDEACANVRVQDSOPEIDNLE 417
 361 YEGHGVRIQDRALINAAQLSARYITGRHLDPKADLVDEACANVRVQDSOPEIDNLE 420
 418 RRMOLIEIHALEREKDKASKARLIEVRKELDLDRDKLOPLTKMYRKEKERIDEIRLK 477
 421 RRIQLEVEIHALEREKDKASKARLIEVRKELDLDRDKLOPLTKMYRKEKERIDEIRLK 480
 478 OKREELMFSLOAERRYDLARAADLYGAIQEVESAIAQLEGTSEENVMLTENVGPHI 537
 481 QRREELQFTLOAERRMDLARVADLYGAIQEVESAIAQLEGTSEENVMLTENVGPHI 539
 538 AEVVSRTWGTIPVTRLGONEKERLIGLADRLHRRVVGONQAVNAVSEAILRSRAGLGRAQ 597
 540 AEVVSRTWGTIPVTRLGONEKERLIGLADRLHRRVVGONQAVNAVSEAILRSRAGLGRAQ 599
 598 PTGSFLFGTGVGKTELAKALAEQFDDENLLVRIDMSEYMEQHSVRLIGAPGYVGH 657
 600 PTGSFLFGTGVGKTELAKALAEQFDDENLLVRIDMSEYMEQHSVRLIGAPGYVGH 659
 658 EGGQLEAVRRPYCVILFDEVEKAHVAFTLLQVLDGRLTDGQRTVDFRNVITM 717
 660 EGGQLEAVRRPYCVILFDEVEKAHVAFTLLQVLDGRLTDGQRTVDFRNVITM 719
 718 TSNLGAHLLAGLTGKVTMEVARDVYRVRHFRPELLNRLDEIVFDPGLSHDQLRKA 777
 720 TSNLGAHLLAGLTGKVTMEVARDVYRVRHFRPELLNRLDEIVFDPGLSHDQLRKA 779
 778 RLQKMDVAVLAERGVALAVTDAALDYILAESVDPVYGARPIRRMMKKVTELSSQVVR 837
 780 RLQKMDVAVLAERGVALAVTDAALDYILAESVDPVYGARPIRRMMKKVTELSSQVVR 839
 838 EETDENSTVYIDAGD-LVYRVE-SGGLVDASTGKSDVLTHIANGP---KRSDAAQA 891
 840 EETDENSTVYIDAGD-LVYRVE-SGGLVDASTGKSDVLTHIANGP---KRSDAAQA 899
 892 VKQWRIEEDDNDNERMIED 911
 900 VKQWRIEEDDNDNERMIED 917
 QSPH4 PRELIMINARY; PRT; 913 AA.
 QSPH4
 01-MAY-2000 (TrEMBLrel. 13, Created)
 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 Heat shock protein 101.
 HSP101.
 Triticum aestivum (Wheat).
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; Pooideae;
 Triticeae; Triticum.
 NCBI TaxID=4565;
 (1)
 SEQUENCE FROM N.A.

RC STRAIN=cv. Mustang;
 RX MEDLINE=21240210; PubMed=11342108;
 RA Campbell J.L., Klueva N.Y., Zheng H.G., Nieto-Sotelo J., Ho T.H.D.,
 NGuyen H.T.;
 RT "Cloning of new members of heat shock protein HSP101 gene family in
 wheat (Triticum aestivum (L.) Moench) inducible by heat, dehydration,
 and ABA";
 RL Biochim. Biophys. Acta 1517:270-277(2001).
 DR EMBL; AF174433; AAFO1280.1; -;
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR003959; AAA ATPase centr.
 DR InterPro; IPR001270; Chaprinin_c1pA/B.
 DR InterPro; IPR004176; Clp_N.
 DR Pfam; PF00004; AAA; 2.
 DR Pfam; PF02861; Clp_N; 2.
 DR PRINTS; PR00300; CUPPROTEASEA.
 DR SMART; SMO0382; AAA; 2.
 DR PROSITE; PS00870; CLPAB 1; 1.
 DR PROSITE; PS00871; CLPAB 2; 1.
 KW ATP-binding.
 SQ SEQUENCE 913 AA; 100897 MW; 867E2EF137C8E7DB CRC64;
 Query Match 84.0%; Score 3822.5; DB 10; Length 913;
 Best Local Similarity 83.3%; Pred. No. 9e-179;
 Matches 761; Conservative 75; Mismatches 69; Indels 9; Gaps 6;
 QY 1 MNPKEPTHKNTTATTAHELAVNAGHAQFTPLHLGALISDPTGIFPQAISAGGEN--A 58
 DB 1 MNPDKFTHKNTTATTAHELAVNAGHAQFTPLHLGALISDPTGIFPQAISAGGEN--A 60
 QY 59 AQSARVINOALKKLPSQSPPPDDIPASSSLIKVIRRAQAQSRGDTHLAVDOLIMGLL 118
 DB 61 AESFERSVAAALKKLPQSPPPDTPVASTSLVAVRAQAQSRGDTHLAVDOLIMGLL 120
 QY 119 EDSQIDLLNEGVATARKVSEVKLRGKEKGVESAGDTNFQALTKYGRDLVEQAGK 178
 DB 121 EDPOISDALKEAGISAARVKAEEVKLRGGDNRRVESASGDTNFQALTKYGRDLVEQAGK 180
 QY 179 DPVIGRDEIRRVVRLGRRRTKNNPVLVIGPGVGTAVVEGLAQRIKVGDPVNSLTDVR 238
 DB 181 DPVIGRDEIRRVVRLGRRRTKNNPVLVIGPGVGTAVVEGLAQRIKVGDPVNSLTDVR 240
 QY 239 ISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEGKVILFIDEIHLVLGAGTEGSMDDAA 298
 DB 241 VALDMGALVAGAKYRGEFEERLKAFLKEVEAEGKVILFIDEIHLVLGAGTEGSMDDAA 300
 QY 299 LFKPMLARGQLRCIGATTLEBYRKVKVEKDAFERRFQOYVVAEVSVPDTISILGLKEK 358
 DB 301 LFKPMLARGQLRCIGATTLEBYRKVKVEKDAFERRFQOYVVAEVSVPDTISILGLKEK 360
 QY 359 EGHGVRIQDRALINAAQLSARYITGRHLDPKADLVDEACANVRVQDSOPEIDNLER 418
 DB 361 EGHGVRIQDRALINAAQLSARYITGRHLDPKADLVDEACANVRVQDSOPEIDNLER 420
 QY 419 KRMOLIEIHALEREKDKASKARLIEVRKELDLDRDKLOPLTKMYRKEKERIDEIRLK 478
 DB 421 KRIQLEVEIHALEREKDKASKARLIEVRKELDLDRDKLOPLTKMYRKEKERIDEIRLK 480
 QY 479 KREELMFSLOAERRYDLARAADLYGAIQEVESAIAQLEGTSEENVMLTENVGPHIA 538
 DB 481 KREELQFTLOAERRMDLARVADLYGAIQEVESAIAQLEGTSEENVMLTENVGPHIA 539
 QY 539 EVSRTWGTIPVTRLGONEKERLIGLADRLHRRVVGONQAVNAVSEAILRSRAGLGRAQ 598
 DB 540 EVSRTWGTIPVTRLGONEKERLIGLADRLHRRVVGONQAVNAVSEAILRSRAGLGRAQ 599
 QY 599 TGSFPLFGTGVGKTELAKALAEQFDDENLLVRIDMSEYMEQHSVRLIGAPGYVGH 658
 DB 600 TGSFPLFGTGVGKTELAKALAEQFDDENLLVRIDMSEYMEQHSVRLIGAPGYVGH 659
 QY 659 EGGQLEAVRRPYCVILFDEVEKAHVAFTLLQVLDGRLTDGQRTVDFRNVITM 718
 DB 660 EGGQLEAVRRPYCVILFDEVEKAHVAFTLLQVLDGRLTDGQRTVDFRNVITM 719

719 SNLGAELHLAGTCKVTMEVARDVCMREVRKHFPELNLRLDEIVFDPPLSHDQLRKVAR 778
 720 SNLGAELHLAGWGN-SMKVARDLVMOEVRHFRPELNLRLDEIVFDPPLSHDQLRKVAR 778
 779 LQMKDVAVRLAERGVALAVTDAALDYILAESYDPVYGARPIRRMWEKVVTELKMWVRE 838
 779 LQMKDVAVRLAERGVALAVTDAALDVLSYDPVYGARPIRRMWEKVVTELKMWVRE 838
 839 EIDENSTVIYIDAGAG--DLVYRVES--GGLVDASTGKSDVLIHIANGPKESSDAQAQVKM 895
 839 EIDENSTVIYIDAAPSDBELTYGVDKHGLVNNARTGHKSDILIQVPSGAVGDDAHAVKMM 898
 896 RI--EETEDDDNEE 907
 899 KIMQDSGEVDDME 912

Q8L614 PRELIMINARY; PRT; 912 AA.
 Q8L614;
 01-OCT-2002 (TrEMBLrel. 22, Created)
 01-MAR-2003 (TrEMBLrel. 22, Last sequence update)
 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 Heat shock protein 101.
 HSP101.
 Oryza sativa (Rice).
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 NCBI_TaxID=4530;
 [1]

SEQUENCE FROM N.A.
 STRAIN=cv. Nipponbare;
 Agarwal M., Sahi C., Young T., Agarwal S., Mitra S., Ganeshan K.,
 Katiyar-Agarwal S., Gallie D.R., Grover A.;
 "Isolation and analysis of HSP101 from rice";
 Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 EMBL: AJ316025; CAC87117.1; -;
 Gramene; Q8L614; -;
 InterPro; IPR003593; AAA_ATPase.
 InterPro; IPR003959; AAA_ATPase centr.
 InterPro; IPR001270; Chaperin_c1pA/B.
 InterPro; IPR004176; Clp_N.
 Pfam; PF00004; AAA_2.
 Pfam; PF02861; Clp_N; 2.
 PRINTS; PR00300; ClpPROTEASEA.
 SMART; SM00382; AAA; 2.
 PROSITE; PS00870; CLPAB_1; 1.
 PROSITE; PS00871; CLPAB_2; 1.
 ATP-binding.
 SEQUENCE 912 AA; 100836 MW; 07EA168AED0605C1 CRC64;

Query Match 83.7%; Score 3807.5; DB 10; Length 912;
 Best Local Similarity 83.2%; Pred. No. 4.9e-178;
 Matches 760; Conservative 77; Mismatches 69; Indels 7; Gaps 5;
 1 MNPEKTHKNTIATAHELAVNAGHAFTPLHLAGALISDPTGIPQAISSAGGENAA- 59
 1 MNPDPNTHKNTNEALVAEHAIEAEGHAQLTPLHLVAALAAADKGGILRQAISSAGGDAGA 60
 60 -QSAERVINQALKPLPSQPPDDIPASSLLKVRRAQAKSGDTHLAVDQLIMGLL 118
 61 PDSFERYVSGALKPLPSQPPDDIPASTALIKVIRRAQAKSGDTHLAVDQLIMGLL 120
 119 EDSQTRDLLNEVGAVTARYKSEVEKLRGKGGKVESASGDTNFQALTYGRDLVEQAGKL 178
 121 EDSLISDCLKEAGVSAARVRAELEKLRGEGKVESASGDTNFQALTYGRDLVEQAGKL 180
 179 DPVIGRDEIRRVILSRRTKNNPVILGEPGVGKTA VVEGLAQRIKVGDPVNSLTDVRL 238
 181 DPVIGRDEIRRVILSRRTKNNPVILGEPGVGKTA VVEGLAQRIKVGDPVNSLTDVRL 240

QY 239 ISLDMGALVAGAKYGEFEERLKSVLKEVEDAGKVVILFIDBIHLVLAGKTEGSDAAN 298
 DB 241 IALDMGALVAGAKYGEFEERLKA V LKEVEEAGKVVILFIDBIHLVLAGRTEGSDAAN 300
 QY 299 LFKPMLARGQLRCIGATTILEEYRKVVEKDAAFERFQOVYVAESPDPDTISILRLGKSKY 358
 DB 301 LFKPMLARGQLRCIGATTILEEYRKVVEKDAAFERFQOVYVAESPDPDTISILRLGKSKY 360
 QY 359 EGHGVRITODRALINAAQSARYITGRHLPDKAIDLVDCAANVRVQLDSQPEEIDNLER 418
 DB 361 EGHGVRITODRALVVAQAQSARYIMGRHLPDKAIDLVDCAANVRVQLDSQPEEIDNLER 420
 QY 419 KRMQLEIELHALEREKDKASKARLEVEKEKDDLDKLOPLTMKYRKEKERIDEIRLKQ 478
 DB 421 KRIQLEVEHALEREKDKASKARLEVEKEKDDLDKLOPLTMKYRKEKERIDEIRLKQ 480
 QY 479 KREELMFSLOEARRDYIARAADLYGAIQEVSAIAOLEGTSSEENVMTENVGPEHIA 538
 DB 481 RREELQFTLQEAERRMDLARVADLYGALQIDVIAKLE-SETGENMLTETVGPQIA 539
 QY 539 EVVSEMTGIPVTRIGONEKERLIGLADRLHVRVVGQAVNAVSEAILRSRAGLGRQOP 598
 DB 540 EVVSEMTGIPVTRIGONDEERLVGLADRLHVRVVGQAVNAVSEAILRSRAGLGRQOP 599
 QY 599 TGSFLFLGPTGVGKTELAKALAEOLFDDENLLVRIDMSEYMEQHSVRLICAPPYVYGE 658
 DB 600 TGSFLFLGPTGVGKTELAKALAEOLFDDENLLVRIDMSEYMEQHSVRLICAPPYVYGE 659
 QY 659 EGGQITEAVRRRYPYCVILFDEVEKARHVAVENTLLQVLDGRLTDGQRTVDFRNSVLIIT 718
 DB 660 EGGQITEQVRRRYPYCVILFDEVEKARHVAVENTLLQVLDGRLTDGQRTVDFRNTVLIIT 719
 QY 719 SNLGAELHLAGTCKVTMEVARDVCMREVRKHFPELNLRLDEIVFDPPLSHDQLRKVAR 778
 DB 720 SNLGAELHLAGWGN-SMKVARDLVMOEVRHFRPELNLRLDEIVFDPPLSHDQLRKVAR 779
 QY 779 LQMKDVAVRLAERGVALAVTDAALDYILAESYDPVYGARPIRRMWEKVVTELKMWVRE 838
 DB 780 LQMKDVAVRLAERGVALAVTDAALDVLSYDPVYGARPIRRMWEKVVTELKMWVRE 839
 QY 839 EIDENSTVIYIDAG--AGDLVYRVES--GGLVDASTGKSDVLIHIANGPKE-SDAAQAVKK 894
 DB 840 EIDENCTVIYIDAAPHKDELAYRVNDRGLVNAETGQKSDILIQVPGAATGSDAAQAVKK 899
 QY 895 MRIEEIEDDNEE 907
 DB 900 MRIMEDEDGMDDE 912

RESULT 8
 Q9ZT12
 ID Q9ZT12 PRELIMINARY; PRT; 918 AA.
 AC Q9ZT12;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE 101 kDa heat shock protein.
 GN HSP101.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP MEDLINE=9903076; PubMed=9784498;
 RX Wells D.R., Tanguay R.L., Le H., Gallie D.R.;
 RA "HSP101 functions as a specific translational regulatory protein whose
 RT activity is regulated by nutrient status";
 RL Genes Dev. 12:3236-3251(1998).
 RN [2]
 RP SEQUENCE FROM N.A.

Gallie D.R., Tanouay R.L., Wells D.R.;
Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AF083344; AAC83689.2; -;
InterPro: IPR003593; AAA_ATPase.
InterPro: IPR003959; AAA_ATPase_Centr.
InterPro: IPR001270; Chaprinin_c1pA/B.
InterPro: IPR004176; Clp_N.
Pfam: PF00004; AAA; 2.
Pfam: PF02861; Clp_N; 2.
PRINTS: PR00300; CLPPROTEASEA.
SMART: SM00382; AAA; 2.
PROSITE: PS00870; CLPAB_1; 1.
PROSITE: PS00871; CLPAB_2; 1.
ATP-binding; Heat shock.

SEQUENCE 918 AA; 101115 MW; 3450BF52A156E200 CRC64;
Query Match 83.3%; Score 3790; DB 10; Length 918;
Best Local Similarity 81.7%; Pred. No. 3.5e-177;
Matches 752; Conservative 86; Mismatches 70; Indels 12; Gaps 7;
1 MNPDKTHKNTIATAHNAVAGHAQPTPLHLAGALISDPTGIFPOAISSAGGENAA- 59
1 MNPDKTHKNTIATAHNAVAGHAQPTPLHLAGALISDPTGIFPOAISSAGGENAA- 59
60 -OSAEVINGALKLPSPSPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQIMGLL 118
61 GDSFVRLAGALKLPSPSPDDIPASSLLIKVIRRAQAQKSGDTHLAVDQIMGLL 120
119 EDSQIRDLINVEGVATRVKSEVEKLKGE-GKVESAGDNTFOALATYGRDLVEQAGK 177
121 EDAQADCLKEAGVSASVRALDKLRGDSNRKVESAFGDTTFQALATYGRDLVEVAGK 180
178 LDPVIGRDEIRRVVRLSRRTKNNPNVLIGPGVGTAVVEGLAQRIKVGDPVNSLTDVR 237
181 LDPVIGRDEIRRVVRLSRRTKNNPNVLIGPGVGTAVVEGLAQRIKVGDPVNSLTDVR 240
238 LISLDMGALVAGAKYRGFEERLKSILKEVEDAEGKVLIFDEHLVLGAKTSGMDAA 297
241 LVALDMGALVAGAKYRGFEERLKSILKEVEDAEGKVLIFDEHLVLGAKTSGMDAA 300
298 NLFKPLARGQLRCIGATTLBEYRYKVEKDAAFERRFOVVVAESPVPDITSLRGLKEK 357
301 NLFKPLARGQLRCIGATTLBEYRYKVEKDAAFERRFOVVVAESPVPDITSLRGLKEK 360
358 YEGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVEACANVRVQLDSQPBEIDNLE 417
361 YEGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVEACANVRVQLDSQPBEIDNLE 420
418 KRMQLELHALEREKOKASKARLIEVRKELDLRLDKLOPLTMKYRKEKERIDEIRLKL 477
421 KRMQLELHALEREKOKASKARLIEVRKELDLRLDKLOPLTMKYRKEKERIDEIRLKL 480
478 KRMQLELHALEREKOKASKARLIEVRKELDLRLDKLOPLTMKYRKEKERIDEIRLKL 537
481 KRMQLELHALEREKOKASKARLIEVRKELDLRLDKLOPLTMKYRKEKERIDEIRLKL 539
538 AEVVSRTGTIPVTRLGQNDKERLIGADRLHKKRVVQGNQAVNAVSEAILSRAGLGRAQQ 597
540 AEVVSRTGTIPVTRLGQNDKERLIGADRLHKKRVVQGNQAVNAVSEAILSRAGLGRAQQ 599
598 PTGSFLFGPTGVGKTELAKALAEQLEDDENLVRIDMSYMEQHSVRLIGAPPGVGH 657
600 PTGSFLFGPTGVGKTELAKALAEQLEDDENLVRIDMSYMEQHSVRLIGAPPGVGH 659
658 EGGQLTEVRRRPPYCVILFDEVEKAHVAFTNLLQVLDGRLTDGQGRVDPFRNVIIM 717
660 EGGQLTEVRRRPPYCVILFDEVEKAHVAFTNLLQVLDGRLTDGQGRVDPFRNVIIM 719
718 TSNLGAHLLAGTCKVTMVARDCVNRVREKFRPELLNRLDEIVVFDPLSHDLQKVA 777
720 TSNLGAHLLAGTCKVTMVARDCVNRVREKFRPELLNRLDEIVVFDPLSHDLQKVA 779
778 RLQMKDVAVRLAERGVAVATDAALDYLLAESDVPYVGPARRRWEKKVVTLSKQVVR 837

Db 780 RLQMKDVAVRLAERGVAVATDAALDYLLAESDVPYVGPARRRWEKKVVTLSKQVVR 839
QY 838 EIDENSTVYIDAGAGD-LVYRVV-SGGLVDASTGKSDVLIHIANGP---KSDRAQA 891
Db 840 EIDENSTVYIDAGAGD-LVYRVV-SGGLVDASTGKSDVLIHIANGP---KSDRAQA 891
QY 892 VKKWRIEIEIEDDDNEEMIED 911
Db 900 VKKWRV--MEDGDEGMD 917

RESULT 9

Q9SYS9 PRELIMINARY; PRT; 582 AA.
AC Q9SYS9
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 101 kDa heat shock protein (Fragment).
GN HSP101.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97134920; PubMed=8980480;
RA Boston R.S., Viitanen P.V., Vierling E.;
RT "Molecular chaperones and protein folding in plants.";
RL Plant Mol. Biol. 33:191-222(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99234099; PubMed=10216257;
RA Nieto-Sotelo J., Kannan K.B., Martinez L.M., Segal C.;
RT "Characterization of a maize heat-shock protein 101 gene, HSP101,
RT encoding a ClpB/Hsp100 protein homologue.";
RL Gene 230:187-195(1999).
DR EMBL: AF083327; AAD26530.1; -;
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR001270; Chaprinin_c1pA/B.
DR PRINTS: PR00300; CLPPROTEASEA.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00871; CLPAB_2; 1.
KW ATP-binding; Heat shock.
FT NON_TER
SQ SEQUENCE 582 AA; 65767 MW; 9ABEAC1FAFE26B7C CRC64;

Query Match 55.0%; Score 2502.5; DB 10; Length 582;
Best Local Similarity 85.2%; Pred. No. 1.6e-114;
Matches 497; Conservative 49; Mismatches 32; Indels 5; Gaps 4;

QY 329 AFERRFOQVVAESPVPDITSLRGLKEKYGHHGVRIQDRALINAAQLSARYITGRHLP 388
Db 1 AFERRFOQVVAESPVPDITSLRGLKEKYGHHGVRIQDRALINAAQLSARYITGRHLP 60
QY 389 DKAILDVEACANVRVQLDSQPBEIDNLERKMQLELHALEREKOKASKARLIEVRKE 448
Db 61 DKAILDVEACANVRVQLDSQPBEIDNLERKMQLELHALEREKOKASKARLIEVRKE 120
QY 449 LDDLKDLQPLTMKYRKEKERIDEIRLKKRBEIMPSLOAEERRYLARAADLRYGAIQ 508
Db 121 LDDLKDLQPLTMKYRKEKERIDEIRLKKRBEIMPSLOAEERRYLARAADLRYGAIQ 180
QY 509 EVESAIAOLEGTSSEENVMLTENVGPEHIAVSRWTGTIPVTRLGQNDKERLIGADRLH 568
Db 181 EIDRAISKLE-SETGENMLTETVGPQIAEVSVSRWTGTIPVTRLGQNDKERLIGADRLH 239
QY 569 KRVVGQONAVNAVSEAILSRAGLGRAQQPTGSFLFGPTGVGKTELAKALAEQLEDDEN 628
Db 240 QRVVGQONAVNAVSEAILSRAGLGRAQQPTGSFLFGPTGVGKTELAKALAEQLEDDEN 299

```

629 LLVRLDMSEYMEQHSVRLIGAPPVGVGHEGGQTEAVRRPYCVILFDEVEKAHVAF 688
300 LLVRLDMSEYMEQHSVRLIGAPPVGVGHEGGQTEAVRRPYCVILFDEVEKAHVAF 359
689 NTLQVLDGRITDGGRTVDPRNSVIMTSLNLAHLLAGLTGKVTNEVARDVCMREVR 748
360 NTLQVLDGRITDGGRTVDPRNTVIMTSLNLAHLLAGLVGKSMKVDRLVMQEV 419
749 KHFRPELNLRLDEIVVFDPLSHDQRLKRVARLQMKDVAVRLAERGVALAVTDAALDYILAE 808
420 RHFRPELNLRLDEIVVFDPLSHDQRLKRVARLQMKDVAVRLAERGVALAVTDAALDYILAE 479
809 SYDPVYGARPIRRWEKKVTVLSKRVVREIDENSTVIYDAGAG--DLVYRVF-SGGLV 865
480 SYDPVYGARPIRRWEKKVTVLSKRVVREIDENSTVIYDAGAG--DLVYRVF-SGGLV 539
866 DASTGCKSDVLIHLIANGPKRDAAQVKKRI-REIEDDDNEE 907
540 NAEIGMCKSDVLIHLIANGPKRDAAQVKKRI-REIEDDDNEE 582

RESULT 10
023323 PRELIMINARY; PRT; 668 AA.
01-JAN-1998 (TrEMBLrel. 05, Created)
01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
Heat shock protein-like protein (Heat shock protein like).
D13375W OR A74G14670.
Arabidopsis thaliana (Mouse-ear cress).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
eurosids II; Brassicales; Brassicaceae; Arabidopsids.
NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
Bevan M., Stiekema W., Murphy G., Wambutt R., Pohl T., Terry N.,
Kreis M., Kavanagh T., Entian K.D., Rieger M., James R.,
Puigdomenech P., Hatzopoulos P., Obermaier B., Duesterhoft A.,
Jones J., Palme K., Ansorge W., Delseny M., Bancroft I., Mewes H.W.,
Schueller C., Chalwatzis N.,
Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
EU Arabidopsis sequencing project;
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
EMBL; Z97336; CAB46061.1; -.
EMBL; AL161539; CAB78509.1; -.
InterPro; IPR0031593; AAA_ATPase.
InterPro; IPR003959; AAA_ATPase_cent.
InterPro; IPR001270; Chaperin_clpA/B.
InterPro; IPR004176; Clp_N.
Pfam; PF00004; AAA; 2.
Pfam; PF02861; Clp_N; 1.
PRINTS; PR00300; CLPPROTEASA.
SMART; SM00382; AAA; 1.
ATP-binding; Heat shock.
SEQUENCE 668 AA; 74043 MW; SEEA7697FC4AC365 CRC64;

Query Match 53.4%; Score 2431; DB 10; Length 668;
Best Local Similarity 70.8%; Pred. No. 6.2e-111;
Matches 499; Conservative 76; Mismatches 90; Indels 40; Gaps 6;

1 MNPEKFTKNTETIATAHELAVNAGHAQFTPLHLAGALISDPTGIFPOAISAG-GENAA 59
1 MNDLRFDPNVKILASARSHAMSLSGQVTPHLGVTLISDLSVTFYRAITSAGDGISA 60
60 QSAERVINALKLSQSPPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLLE 119
61 QSVNVINQSLYK-----TKRNLGDTKGVAVLVLSILE 95
120 DSQIRDLNLNEGVATARKVEKVEKLRGKGVESASGDTNFQALKTYGRDLVQAGKLD 179

```

```

Db 96 DSQISDVLKAGVVPKVKSEVEKLR-----GEVILFAALKTYGTDLVQAGKLD 144
Qy 180 PVIGRDEIRRVRIILSRRTKNNPVLIIGPGVGTAVVEGLAQRIKVGDPVNSLTVDRLI 239
Db 145 PVIGRDEIRRVRIILSRRTKNNPVLIIGPGVGTAVVEGLAQRIKVGDPVNSLTVDRLI 204
Qy 240 SLDMGALVAGAKYRGFEERLKSVLKEVEDAGKVLFIIDEIHLVLGAGKTGSGMDAANL 299
Db 205 SLDFGAMVAGTTLRGQFEERLKSVLKAVEAQKVVLFIDEIHMALGACKASGSDAAKL 264
Qy 300 FKPMALARGQLRCIGATTLEBYKYVEKDAAPERRRQOVYVAEPSPDPTIILRGLKEKYE 359
Db 265 LKPMALARGQLRFIGATTLEBYKYVEKDAAPERRRQOVYVAEPSPDPTIILRGLKEKYE 324
Qy 360 GHGVRIQDRALINAQLSARYITGRHLPLDKAIDLVDBACANVRVQLDSQPEIDNLRK 419
Db 325 GHGVRIQDRALVLSAQLSERVITGRRLPDKAIDLVDBACANVRVQLDSQPEIDNLRK 384
Qy 420 RMQLEIEHLALEREK-DKASKARLIEVRKELDDLRLKLOPLTKYRKEKERIDEIRRLKQ 478
Db 385 VMQLEIEHLALEREK-DKASKARLIEVRKELDDLRLKLOPLTKYRKEKERIDEIRRLKQ 444
Qy 479 KREELMFSLOEAEERYDLARAADLVYGAIOEVESAIAOLEGTSSEENVMLTENVGPEHIA 538
Db 445 NEDDLMLIALQEAERORHVPKAAVLKYGAIQEVESAIAKLE-KSAKONVMLTETVGPENIA 503
Qy 539 EVVSRWTGIPVTRLQNEKRLIGLADRLHRRVVGQNOAVNAVSEAILRSRAGLGRQOP 598
Db 504 EVVSRWTGIPVTRLQNEKRLIGLADRLHRRVVGQNOAVNAVSEAILRSRAGLGRQOP 563
Qy 599 TGSFLFGPTGVGKTELAKALAEQFDENLLVRIDMSYNEQHSVRLIGAPPG-YVGH 657
Db 564 TGSFLFGPTGVGKTELAKALAEQFDENLLVRIDMSYNEQHSVRLIGAPPG-YVGH 623
Qy 658 REGQLTEAVRRPYCVILFDEVEKAHVAVENTLQVLDGRITD 702
Db 624 REGQLTEAVRRPYCVILFDEVEKAHVAVENTLQVLDGRITD 668

RESULT 11
QBDJ40
ID Q8DJ40 PRELIMINARY; PRT; 871 AA.
AC Q8DJ40;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ClpB protein.
GN TLR1389.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RL Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002);
DR EMBL; AP005373; BAC08941.1; -.
KW Complete proteome.
SQ SEQUENCE 871 AA; 98389 MW; D19D08E1A680F5FB CRC64;

Query Match 50.5%; Score 2299.5; DB 16; Length 871;
Best Local Similarity 53.1%; Pred. No. 2.4e-104;
Matches 456; Conservative 177; Mismatches 209; Indels 17; Gaps 9;

Qy 2 NPEKFTKNTETIATAHELAVNAGHAQFTPLHLAGALISDPTGIFPOAISAGGENAAQS 61

```

5 NPQTEKAWAAIAITPDLAKQAQHONLESHLMKSL--E0GLATQIFQAG--CSVQR 61
 62 AERVINQAKKLPSPSPDDIPASSSLIKVIRRAQAQKRGDTLHVAQDQIMGLLEDS 121
 62 IRDLTDFEITSQPKIS--HPSGYLQSLDKLLDRAEARKQGFDEFISIEHLVLAFAQDD 120
 122 QI--RDLLNEVGATARKVSEKELRGKGGKVESAGDTNFQALKTYGRD--LVEQAGKL 178
 121 RFGKGLFDQIGLSEKVLBEATQIQNGSQ--KYTDQNPGEKTAALAEKYGRDITLLARQOKL 178
 179 DPVIGRDEIRRVIRILSRRTKNNPVLIGEPGVGTAVVEGLAQRIKVGDPVNSITDVRIL 238
 179 DPVIGRDEIRRVIRILSRRTKNNPVLIGEPGVGTAVVEGLAQRIKVGDPVNSITDVRIL 238
 239 ISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEGKVLIFIDEIHLVLGAGKTEGSMADAAN 298
 239 IALDMGALVAGAKYRGEFEERLKSVLKEVEDAEGKVLIFIDEIHLVLGAGKTEGSMADAAN 298
 299 LFKPMLARGOLRCIGATTLEVRKYVEKDAAFERRFOQYVYABSPVDTTILRLGKEKY 359
 299 LFKPMLARGOLRCIGATTLEVRKYVEKDAAFERRFOQYVYABSPVDTTILRLGKEKY 359
 359 EGHGVRIDQDALINAAQSARYITGRHLDPKALDLDVEACANVRVOLDSDOPEIDNLER 418
 359 EIHGKVISDTALVAAATLSARYISDRFLPKADLDLVEAANKLMEITSXPELDEIDR 418
 419 KRMQLELHLEHREKOKASKARLIEVRKELDDLRDKLQPLTMKYRKEKERIDIRRLKQ 478
 419 KILQJEMERLSQKETSAAASDRLEKLELRELADLKEQSRNLNAQWAEKEVIDRLQSIKE 478
 479 KREELMSLOAERRYDLARAADIRYGAIOVESAIQALETSSSENV----MLTENVP 534
 479 EIEKNVIEIOAERNYDLNRAAEKLYGKLTLEHKLAEAEAKLEIQVGGSRSLRDEVTE 538
 535 EHIAEVSRSWTGIPVTRILGONEKERLIGLADRLHKKRVVGQNOAVNAVSEAILRSRAGLGR 594
 539 ADIAEILSKWTGIPVSKLVESEAQLHLEBELHKKRVVGQDEAVSAVAEAIQSRAGLAD 598
 595 AQQTGSPFLFGPTGVGKTELAKALAEQFDENLVRIDMSEYMEQHSVRLTGAPPY 654
 599 PNRPIASFIFGPTGVGKTELAKALAEQFDENLVRIDMSEYMEQHSVRLTGAPPY 658
 655 VGEEGGOLTEAVERRPVCLVLFDEVEKAHVAVENTLLOVLDGRLTDGGRTVDFRNSV 714
 659 VGYDEGGOLTEAVERRPVCLVLFDEVEKAHVAVENTLLOVLDGRLTDGGRTVDFRNSV 714
 715 IIMTSNLGAEHL--AGLTGKVTMEVARDCVMRVRKHFRPELNLRLDEIVVDFPLSHDQ 772
 719 IIMTSNLGAEHL--AGLTGKVTMEVARDCVMRVRKHFRPELNLRLDEIVVDFPLSHDQ 776
 773 LRKVARLQMDVAVRVAERGVAVLAVTDAALDYILAESYDPVYGARPIRMWKKVVTLS 832
 777 LRQIVQLQVORLQORLSDRHITLSLTKAIDFLAEVGYDPVYGARPLKRAIQOLETPIA 836
 833 KVVVREIDENSTVIIDAG 851
 837 KAILRGDFDGDITLVDVG 855

JUL 12 2004

000828 PRELIMINARY; PRT; 867 AA.
 01-JUL-1997 (TrEMBLrel. 04, Created)
 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 Heat shock protein 100.
 HSP100.
 Leishmania donovani.
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 NCBI_TaxID=5661;
 [1]
 SEQUENCE FROM N.A.

RC STRAIN=MHOM/IN/72/DD8;
 RA Clos J.;
 RT "Amastigote stage-specific expression of the Leishmania donovani heat
 RT shock protein 100.";
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z94053; CAB08073.1; -
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR001395; AldG/ket_red
 DR InterPro; IPR001270; Chaprinin_clpA/B.
 DR InterPro; IPR004176; Clp_N.
 DR Pfam; PF00004; AAA; 1.
 DR Pfam; PF02861; Clp_N; 1.
 DR PRINTS; PR00300; CLP_PTEASEA.
 DR SMART; SM00382; AAA; 2.
 DR PROSITE; PS00063; ALDOXETO_REDUCTASE_3; 1.
 DR PROSITE; PS00870; CLPAB_1; 1.
 DR PROSITE; PS00871; CLPAB_2; 1.
 KW ATP-binding.
 SQ SEQUENCE 867 AA; 97063 MW; E8B221E860A26DB2 CRC64;
 Query Match 50.5%; Score 2298; DB 5; Length 867;
 Best Local Similarity 54.4%; Pred. No. 2.8e-104;
 Matches 462; Conservative 153; Mismatches 224; Indels 10; Gaps 5;
 QY 3 PEKFTHTKMTETATATAHELAVNAGHAQFTPLHLAGALISDPTGIFPQAISAGGENAAQSA 62
 DB 6 PE-WTQAASDLWARTAAALARKKANGYLDVPVHLAYVMFEDENSLASRVVKLG---AAV 60
 QY 63 ERVINQAKKLPSPSPDDIPASSSLIKVIRRAQAQKRGDTLHVAQDQIMGLLEDSQ 122
 DB 61 KDLGARVDALPTQMPAPTQPRPNSDMRMVMTAQERVALGDITLMAADHFLALHESKE 120
 QY 123 IRDLNEVGATARKVSEKELRGKGGKVESAGDTNFQALKTYGRDLVEQA--GKLD 180
 DB 121 VGRILDAAGAKKARATILLEM--KGKINSDFODDNYESLNKYAVDLCKQAEKGKLD 178
 QY 181 VIGRDEIRRVIRILSRRTKNNPVLIGEPGVGTAVVEGLAQRIKVGDPVNSITDVRILIS 240
 DB 179 VIGRADEILTRVLSRRTKNNPVLIGEPGVGTAVVEGLAQRIKVGDPVNSITDVRILIS 238
 QY 241 LMGALVAGAKYRGEFEERLKSVLKEVEDAEGKVLIFIDEIHLVLGAGKTEGSMADAANLF 300
 DB 239 LMGALVAGAKYRGEFEERLKSVLKEVEDAEGKVLIFIDEIHLVLGAGKTEGSMADAANLF 298
 QY 301 KPMARGOLRCIGATTLEVRKYVEKDAAFERRFOQYVYABSPVDTTILRLGKEKTEG 360
 DB 299 KPMARGOLRCIGATTLEVRKYVEKDAAFERRFOQYVYABSPVDTTILRLGKEKTEG 358
 QY 361 HHGVRIDQDALINAAQSARYITGRHLDPKALDLDVEACANVRVOLDSDOPEIDNLERKR 420
 DB 359 HHGVRIDQDALINAAQSARYITGRHLDPKALDLDVEACANVRVOLDSDOPEIDNLERKR 418
 QY 421 MQLETELHLEHREKOKASKARLIEVRKELDDLRDKLQPLTMKYRKEKERIDIRRLKOKR 480
 DB 419 MQLETELHLEHREKOKASKARLIEVRKELDDLRDKLQPLTMKYRKEKERIDIRRLKOKR 478
 QY 481 EELMFSLOEAEERYDLARAADIRYGAIOVESAIQALETSSSENVMLTENVPGEHIAE 539
 DB 479 EELMFSLOEAEERYDLARAADIRYGAIOVESAIQALETSSSENVMLTENVPGEHIAE 538
 QY 540 VVSRWTGIPVTRILGONEKERLIGLADRLHKKRVVGQNOAVNAVSEAILRSRAGLGRAOPT 599
 DB 539 VVSRWTGIPVTRILGONEKERLIGLADRLHKKRVVGQNOAVNAVSEAILRSRAGLGRAOPT 598
 QY 600 GSFLFLGPTGVGKTELAKALAEQFDENLVRIDMSEYMEQHSVRLTGAPPYGVGHEE 659
 DB 599 GSFLFLGPTGVGKTELAKALAEQFDENLVRIDMSEYMEQHSVRLTGAPPYGVGHEE 658
 QY 660 GGOLTEAVERRPVCLVLFDEVEKAHVAVENTLLOVLDGRLTDGGRTVDFRNSVIMTS 719
 DB 659 GGOLTEAVERRPVCLVLFDEVEKAHVAVENTLLOVLDGRLTDGGRTVDFRNSVIMTS 718

720 NLGAHLLAGLTGKVTMEVARDVNRVVRKHFRPELNLRLDEIVVDFPLSHDQRLKRVARL 779
 719 NLGAQYQNDMTSPKAVEYVACTQVMEVGEVKEFRPEFNLRLDDILFRSLGLKEMTGIDL 778
 780 QMKDVAVRABRGVALAVTDAAALDYVILAEVDVYVVGARPIRRWMEKVVVLSKVVVREE 839
 779 ITEELYGLKXQDSRVSLEAKYVLESADFADMGARPLRRVWEKNITTELSEMTISOE 838
 840 IDENSTVVI 848
 839 LSPNSTVKV 847

SUULT 13
 Q8YJ91 PRELIMINARY; PRT; 931 AA.
 01-MAR-2002 (TrEMBLrel. 20, Created)
 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 ATP-dependent Clp protease, ATP-binding subunit CLPB.
 BMEI0195.
 Brucella melitensis.
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Brucellaceae; Brucella.
 NCBI_TaxID=29459;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=16M / ATCC 23456 / Biotype 1;
 MEDLINE=20020109; PubMed=11756688;
 DeVecchio V.G., Kapural V., Redkar R.J., Patra G., Mujer C., Los T.,
 Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
 Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 Haselkorn R., Kyrides N., Overbeek R.;
 "The genome sequence of the facultative intracellular pathogen
 Brucella melitensis";
 Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
 EMBL; A8009461; ALU51377.1; -
 InterPro; IPR003593; AAA ATPase.
 InterPro; IPR003959; AAA ATPase centr.
 InterPro; IPR001270; Chaprinin-clpA/B.
 InterPro; IPR004176; Clp_N.
 Pfam; PF00004; AAA; 1.
 Pfam; PF02861; Clp_N; 2.
 PRINTS; PR00300; CLPPTASEA.
 SMART; SM00382; AAA; 2.
 PROSITE; PS00870; CLPAB_1; 1.
 PROSITE; PS00871; CLPAB_2; 1.
 Complete proteome.
 SEQUENCE 931 AA; 103217 MW; C227C61BC13AC076 CRC64;

Query Match 50.4%; Score 2294; DB 16; Length 931;
 Best Local Similarity 53.5%; Pred. No. 4.9e-104;
 Matches 471; Conservative 151; Mismatches 244; Indels 14; Gaps 7;

1 MNPEKFTKNTETATATHELAVNAGHPTPLHAGALISDPTGIFPOAISSAGGENAAQ 60
 58 MNIEKTVRVRGFIQSAQTFALSSGNOQFTPEHLKVLIDDDDEGLAASLVERAGR--VG 115
 61 SAERVINOAKKLPSQPPDDIPASSLSLKVIREAQAQKSRGDTLAVDQLTMGLL-- 118
 116 DVMRGLOSALEKLPKVSNGNDQLYLSQPLAKVFLSABELASKAGDSFVTVRRLTALAME 175
 119 EDSQTRDLINVEGVATARKVSEKLRGKGGKKBESAGDTNFQALTKYGRDLVE--QAG 176
 176 KSAKTSSEILSAGVTPALNRVINDMR--KGRADSASAESNYDALKKYARDLTEDARAG 233
 177 KLDPVIGDEEIRRVRLSRKNPVLIGEPGVGTAVVEGLAQRIKVGDPVNSLTDV 236
 234 KLDPVIGDEEIRRTIQVLSRKNPVLIGEPGVGTAVVEGLAQRIKVGDPVNSLTDV 293
 237 RLISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEKGVLIFDIHLVGLNGKTEGSDA 296

Db 294 QLMALDMGALVAGAKYRGEFEERLKSVLKEVEDAEKGVLIFDIHLVGLNGKTEGSDA 353
 QY 297 ANLFKPLMARGLCIGATTLEERYKYVEKDAFFERRFOQVVAEPSPVDTTISIILRGK 356
 Db 354 SNLLKPALARGELHCVGATTLLEERYKYVEKDAALARRPQVFDVETVEDTISIILRG 413
 QY 357 KYEGHHGVRIQDRALINAQLSARIYTORHHPDKAIDLVDACANVRVOLDSPQPEIDNL 416
 Db 414 KYEQHHKVRVSDSALVAAALSNRYITDRFLPKAIDLVDAAASRLRMQVDSKPELDEI 473
 QY 417 ERKRMQLEIHELHAREKDKASKARLIEVRKELDDLRDLQPLTKMYKEKERIDIEIRL 476
 Db 474 DERIMQLKIERALKVETDAASKORLQRIEKLSELSDEESAEKTAQWAEKOKLGLA 533
 QY 477 KQKRELMFSIQEARRRYDLARAADRLYGAIQEVESATAQLEGTSSEEN--VMLTENVP 534
 Db 534 KQLEBEARNALAIQORNGEPQKAGELAYGTIPQLEKQLADAE--SQENKGLSEETVTP 590
 QY 535 EHIAEVRVSWTGPVTRLGQNEKERLGLADRLHVRVVGQNAVSEAILRSRAGLGR 594
 Db 591 DRVAQVISRWGTGIPVDRMLEGEREKLEMEDEIGKRVVGQGEAVQAISKAVRRARAGLQD 650
 QY 595 AQQPTGSEFLGPTGVGKTELAKALAEQLPDENLVRIDMSEYMEQHSYRLIGAPPGY 654
 Db 651 PNRPIGSPFLGPTGVGKTELTKALASFLFODDTAMVRIDMSEFMKHSYRLIGAPPGY 710
 QY 655 VGHGEGGOLTEAVRRPYCVILFDEVEKARHVAVENTLLQVLDDGRLTDGQRTVDRNSV 714
 Db 711 VGYBEGGVLTEAVRRPYQVILFDEIEKAPDVFNVLLQVLDDGRLTDGQHTVDFNTV 770
 QY 715 IIMTSNLSGAEHLLAGLTGKVTMEVARDVNRVVRKHFRPELNLRLDEIVVDFPLSHDQ 774
 Db 771 IIMTSNLSGAEYLV-NLGENDDVETVRDVMGVVRAVSPRPEFLNRVDEIILFHLRREDMG 829
 QY 775 KVARLQMKDVAVRABRGVALAVTDAAALDYVILAEVDVYVVGARPIRRWMEKVVVLSK 834
 Db 830 AIVDIQMRQLYLLSDRKITVLEDDAREWLANKGDPAYGARPLKRVIOKEVQDPLAER 889
 QY 835 VREIDENSTVIYIDAGADLVVRVSGGLVDASTGKSD 874
 Db 890 ILLGDLGSLVKITAGSDRLNFRPISGAFSAAPERDE 929

RESULT 14
 Q9AEM5
 ID Q9AEM5 PRELIMINARY; PRT; 874 AA.
 AC Q9AEM5;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE ClpB protein.
 GN CLPB OR B1864.
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 CX NCBI_TaxID=29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21172887; PubMed=11274130;
 RA Ekaza E., Teyssier J., Ouahrani-Bettache S., Liautard J.P.,
 RA Koehler S.;
 RT "Characterization of Brucella suis clpB and clpAB mutants and
 RL participation of the genes in stress responses";
 RN J. Bacteriol. 183:2677-2681 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,

[illegible]

239 LDMGALIAGAKYGEPEERLKAVALNEVKSNDKILFIVEIHVLGAGKSGDAMDAANLL 298
301 KPMIARGQLRCIGATTLEEYKVEKDAAFERRFOQVYVAEPSVPTTISILRGLKEKYEG 360
299 KPLLARGDVTTIGATTLEEYKVEKDAAFERRFMPVQVHEPVEBECTSILRGLKDRYEQ 358
361 HGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQDQSPHIDNLERKR 420
359 HGVQITDKAVVAAQLAGRYITNRFLPKAIDLIDEACANVRVTLLSSRPABIDALERKK 418
421 MOLEIELHALEREKDKASKARLIEVEKELDDLDKLOPLTMKYRKEKERIDEIRLKQKR 480
419 RQLEIEERALQORDKASAKERLKAVERKAEIOKVEEKLGLLAKYQERGRIDELQATQAKL 478
481 EELMFSLQBAERYDLARAADLRYGAIQEVESAIAQL-EGTSGSEENVMLTENVGPEHIAE 539
479 DEKKVKLEAERMOMETAADLUKYVIPILQDRIRSLKEAIEKQKATMLQGTVTGTDIAT 538
540 VVSRWTGIPVTRIGONEKERLIGLADRLHRRVVGQNAVNSHAILRSRAGLGRAOQPT 599
539 VVSRWTNIPVTKLSOTERERLLHLADQLHLRVKQDEAVSRVAEAILRSRAGLARSDRPT 598
600 GSFLFLGPTGVGKTELAKALAEOLFDDENLVRIDMSEYMECHSYSLIGAPPGYVGHEE 659
599 GSFLFLGPTGVGKTELSKAVASELFDKATMVKLDMSEYMECHSYSLIGAPPGYVGHEE 658
660 GGQTEAVRRRPPYCVILFDEVEKAHVAVFNTLLQVLDGRLTDGQRTVDFRNSVIIMTS 719
659 GGQTEPVRPPYVTVLLDEVEKAHFNFNVLQVLDGRLTDSHGRTVDFCFTIIMTS 718
720 NLGAHELLAGLTKVTVMEVARDCVMEVRKFRPELLNRLDEIVVDFDPLSHQLRKVARL 779
719 NLGAQYLNQNDTSPKPEYVAQAQVMEVGRKFRPEFINRLDDIILFRSLGKEMTGII DL 778
780 QMKDVAVRLAERGVAVLDAALDYILAESYDPVYCARPIRWMKKVVTTELSKMVRREE 839
779 ITEELNRLXQDSIRVSLTEAKQVLESAPFADMGARPIRWRVEXNITTELSRMIISQE 838
840 IDENSTVYI 848
839 LSPNSTVKV 847

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

Protein - protein search, using sw model

February 13, 2004, 01:28:20 ; Search time 45 Seconds
(without alignments)
856.560 Million cell updates/sec

US-09-812-350-17
Select score: 4550
Sequence: 1 MNPEKTHKNETIATAHEL.....VKMKRIEIEDDDNEEMIED 911

Index table: BLOSUM62
Gapop 10.0 , Gapext 0.5

328717 seqs, 42310858 residues

328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

--processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
 - 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
 - 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
 - 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
 - 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pap.*
 - 6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2194	48.2	857	4	US-08-887-534A-47
2	2194	48.2	857	4	US-09-227-431-47
3	2173.5	47.8	875	4	US-09-107-532A-4537
4	2140	47.0	863	4	US-09-328-352-6730
5	2133	46.9	859	4	US-09-199-637A-281
6	2132	46.9	859	4	US-09-252-991A-21413
7	2108	46.3	871	4	US-09-134-001C-3979
8	1912	42.0	872	4	US-09-138-452A-163
9	1839	40.4	908	2	US-08-249-380-2
10	1830	40.2	823	4	US-09-134-001C-4081
11	1810.5	39.8	866	3	US-09-040-843-2
12	1810.5	39.8	866	4	US-09-621-855-2
13	1789.5	39.3	1044	4	US-09-252-991A-22493
14	1745.5	38.4	672	3	US-09-040-843-4
15	1745.5	38.4	672	4	US-09-621-855-4
16	1682.5	37.0	830	4	US-09-107-532A-5709
17	1604	35.3	845	4	US-09-198-452A-458
18	1450.5	31.9	558	4	US-09-724-623-80
19	1427.5	31.4	745	4	US-09-107-532A-5687
20	1396	30.7	879	4	US-09-252-991A-22604
21	1388.5	30.5	641	4	US-09-724-623-75
22	1387	30.5	1049	4	US-09-252-991A-17298
23	1333.5	29.3	701	3	US-08-923-511-2
24	1333.5	29.3	701	4	US-09-416-874A-2
25	1327.5	29.2	767	4	US-09-252-991A-31198
26	1325.5	29.1	761	4	US-09-328-352-7492
27	1271	27.9	903	4	US-09-328-352-6479

Sequence 25547, A
Sequence 15, Appl
Sequence 31500, A
Sequence 25328, A
Sequence 28426, A
Sequence 14, Appl
Sequence 390, App
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 1, Appl
Sequence 4, Appl
Sequence 32259, A
Sequence 2, Appl
Sequence 3, Appl

US-09-252-991A-25547
US-09-218-363-15
US-09-252-991A-31500
US-09-252-991A-25328
US-09-252-991A-28426
US-09-218-363-14
US-08-858-207A-390
US-08-466-390-4
US-08-470-950-4
US-08-467-781-4
US-08-195-487-4
US-08-483-924-4
US-09-452-294-1
PCT-US93-06160-4
US-09-741-150-4
US-09-252-991A-32259
US-09-592-054-2
US-08-938-105-3

ALIGNMENTS

RESULT 1

US-08-887-534A-47
; Sequence 47, Application US/08887534A
; Patent No. 6455323
; GENERAL INFORMATION:
; APPLICANT: Holder, David W.
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,534A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 28341/33996
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: (312) 474-6600
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 857 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-887-534A-47

Query Match 48.2%; Score 2194; DB 4; Length 857;
Best Local Similarity 51.7%; Pred. No. 8e-175;
Matches 447; Conservative 154; Mismatches 239; Indels 24; Gaps 9;
Qy 1 MNPEKTHKNETIATAHELAVNAGHAQFTPLHAGALISDPTGIPQAISSAGGNAQA 60
Db 1 MRDLINKQLALADAQSLALGHNDQFTEPLHAGALLNQGSGVSPLTSA-GINAGQ 59
Qy 61 SAERVINQALKLPSPQPPDDTPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGL 120

Fri Feb 13 10:59:40 2004

```

60 -LRTDINQALNLPQVEGTGGVQSDQLVRVNLCDNVAQKRGDNFISSELFVLAALLES 118
121 -SQIRDLLNEVGAVATARKVSEVEKLRGKGGKVESASGDTNFQALTKYGRDLVEQA--GK 177
119 RGTVDILKAAGATTANITQATEQMRG--GESVNDQGAEDQKALQKKTIDLTERAEQK 176
178 LDPVIGRDEIRVRVILSRRTKNNPVLIGPGVGTAVVEGLAQIRVKGDPVNSLTDVR 237
177 LDPVIGRDEIRVIRTIQVLRRTKNNPVLIGPGVGTAVVEGLAQIRVKGDPVNSLTDVR 236
238 LISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEKGVILFIDEIHLVLAGKTEGSDMAA 297
237 VLALDMGALVAGAKYRGEFEERLKSVLKEVEDAEKGVILFIDEIHLVLAGKTEGSDMAA 296
298 NLFKPMALRGQLRCIGATTLEEVKRVKVEKDAFERRFQOVVVAEPSVPTTISILRLKKEK 357
297 NMLKPALARGELHCVGATTIDEVYRQVIEKDALEFRFQOVVVAEPSVPTTISILRLKKEK 356
358 YEGHHGVRIQDRALINAAQLSARYITGRHLPKADILVDEACANVRVQLDSQPEEDNLE 417
357 YELHHVQITDPAIVAAATLSHRYIADRLQPKADILVDEACANVRVQLDSQPEEDNLE 416
418 RKMQLIEIHLALEREKDKASKARLIEVRKELDDRLDKLQPLTMKVKKEKIDEIRLKL 477
417 RRIIQLKLEQALMKESDEASKKRLDMLNEELSDKEROYSELEEWKAEKASLSGTQTIK 476
478 QKRELMFSLQEAERRYDILARAADLYRGATQVE---SATAQLEGTSSEENVMLTENVG 534
477 AELEQAKIAIEQARRVGLDARMSELQYKIPLEKQLEAATQLEGKTM---LLRNKVT 533
535 EHAIEVVSRTGIPVTRLGONKEKRLIGLADRLHVRVQGNQAVNVAEAILRSRAGLGR 594
534 AEIAEVLARWTGIPVSRMSESEKRLRMEQELHHRVIGQNEAVDAVSNAIRSRLAGLAD 593
595 AQOQTGFLFGPTGVGKTELAKALAEQLEFDDENLVRIDMSEYMEQHSVSRILGAPPGY 654
594 PNRPIGSLFLGPTGVGKTELCKALANFMFDSDEAMVRIDMSEYMEKHSVSRILGAPPGY 653
655 VGHGEGQLTAVRRRRCYVILFVEVEKAVAVNTLQVLDGRLTDGQRTVDPRNSV 714
654 VGYEGGYLTAVERRRRPSVILDEVEKAPDVFNILLQVLDGRLTDGQRTVDPRNTV 713
715 LIMSNTGABHLAAGLTKGVMEVARDQVMEVRKFRPELNLNRLDRIVDPDPLSHDQLR 774
714 VIMTSLGSD-LIQERFELDYAHKELVLGVVSHNFRPEFINRIDEVVFHPLGEQHIA 772
775 KVARLQMDKAVRLAERGVAVLADTDAALDYILAESYDPVYGARIRRMMEKVTLSKM 834
773 STAQIQLKRLYKRLBERGVEIHSIDEALKJLSENGYPVYGARPLKRAIQOOIENPLAQ 832
835 -----VVREIDENSTVVI 848
833 ILSGELVPGKVRIRVEDNDRIVAV 856

```

RESULT 2
 S-09-527-431-47
 Sequence 47, Application US/09527431
 Patent No. 6485899
 GENERAL INFORMATION:
 APPLICANT: Holden, David W.
 TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 STREET: 233 South Wacker Drive/6300 Sears Tower
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/527,431
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: US/08/887,534
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 28341/33996
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-6600
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 857 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-527-431-47

Query Match 48.2%; Score 2194; DB 4; Length 857;
Best Local Similarity 51.7%; Pred. No. 8e-175;
Matches 447; Conservative 154; Mismatches 239; Indels 24; Gaps 9;

QY 1 MNPEFTHTKNTTATTAHELAVNAGHAQFTPLHLAALISDPTGIFFOAISSAGENAAQ 60
DB 1 MRLDLTKNFQLALADAQSLALGHNDNQFIEPLHMSALLNQEGSVSPLLTSA-GINAGQ 59
QY 61 SAERVINQALKKLPSQSPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLLE 120
DB 60 -LRTDINQALNLPQVEGTGGVQSDQLVRVNLCDNVAQKRGDNFISSELFVLAALLES 118
QY 121 -SQIRDLLNEVGAVATARKVSEVEKLRGKGGKVESASGDTNFQALTKYGRDLVEQA--GK 177
DB 119 RGTVDILKAAGATTANITQATEQMRG--GESVNDQGAEDQKALQKKTIDLTERAEQK 176
QY 178 LDPVIGRDEIRVRVILSRRTKNNPVLIGPGVGTAVVEGLAQIRVKGDPVNSLTDVR 237
DB 177 LDPVIGRDEIRVIRTIQVLRRTKNNPVLIGPGVGTAVVEGLAQIRVKGDPVNSLTDVR 236
QY 238 LISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEKGVILFIDEIHLVLAGKTEGSDMAA 297
DB 237 VLALDMGALVAGAKYRGEFEERLKSVLKEVEDAEKGVILFIDEIHLVLAGKTEGSDMAA 296
QY 298 NLFKPMALRGQLRCIGATTLEEVKRVKVEKDAFERRFQOVVVAEPSVPTTISILRLKKEK 357
DB 297 NMLKPALARGELHCVGATTIDEVYRQVIEKDALEFRFQOVVVAEPSVPTTISILRLKKEK 356
QY 358 YEGHHGVRIQDRALINAAQLSARYITGRHLPKADILVDEACANVRVQLDSQPEEDNLE 417
DB 357 YELHHVQITDPAIVAAATLSHRYIADRLQPKADILVDEACANVRVQLDSQPEEDNLE 416
QY 418 RKMQLIEIHLALEREKDKASKARLIEVRKELDDRLDKLQPLTMKVKKEKIDEIRLKL 477
DB 417 RRIIQLKLEQALMKESDEASKKRLDMLNEELSDKEROYSELEEWKAEKASLSGTQTIK 476
QY 478 QKRELMFSLQEAERRYDILARAADLYRGATQVE---SATAQLEGTSSEENVMLTENVG 534
DB 477 AELEQAKIAIEQARRVGLDARMSELQYKIPLEKQLEAATQLEGKTM---LLRNKVT 533
QY 535 EHAIEVVSRTGIPVTRLGONKEKRLIGLADRLHVRVQGNQAVNVAEAILRSRAGLGR 594
DB 534 AEIAEVLARWTGIPVSRMSESEKRLRMEQELHHRVIGQNEAVDAVSNAIRSRLAGLAD 593
QY 595 AQOQTGFLFGPTGVGKTELAKALAEQLEFDDENLVRIDMSEYMEQHSVSRILGAPPGY 654
DB 594 PNRPIGSLFLGPTGVGKTELCKALANFMFDSDEAMVRIDMSEYMEKHSVSRILGAPPGY 653

```


LENGTH: 863
 TYPE: PRT
 ORGANISM: Acinetobacter baumannii
 09-328-352-6730
 Query Match 47.0%; Score 2140; DB 4; Length 863;
 Best Local Similarity 49.3%; Pred. No. 2.7e-170;
 Matches 424; Conservative 181; Mismatches 245; Indels 10; Gaps 6;
 1 MNPEKTHKNTIATATHELANVAGHAQFTPLHLAGALISDPTGIFPOAISSAGGENAAQ 60
 5 MRFEKFTNRLOALSDAOLAMGKHDTATAGIHILSTLLEPSNI--SLLOQAGAR--LP 60
 61 SAERVINOALKPLSOPSPDDIPASSSLIKVIRRAQAQKSRGDTLHVLQIMGLLE 120
 61 ELKQKLEQALKDAPTIANGTGVNLPNPAKALNADRYAQAGDEFISTDWLLGLAET 120
 121 SQIRDLNNEVGATARKVSEKLGKGGKVESAGDTNFQALNTYGRDLVQEA--GKL 178
 121 GETKNLSAVGTPDLSLRKVIENIRGSD--KVMNNHEDQDRLNKNTIDLTERALSGKL 178
 179 DPVIGRDEIRVRILSRRTKKNPVLLIGEPGVGTAVVEGLAQRIVKGVDPNSLTDVRL 238
 179 DPVIGRDEIRRTIQVLSRRTKKNPVLLIGEPGVGTAVVEGLAQRIVKGVDPNSLTDVRL 238
 239 ISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEAGKVLFIIDELHVLGAGKTEGSDAAN 298
 239 LSLSLGLSLAGAKYRGEFEERLKSVLKEVEDAEAGKVLFIIDELHVLGAGKTEGSDAAN 298
 299 LKPMPLARGOLRCIGATTILEYRKYVEKDAAFERRPQQVYVAEPSVDPDTISILRGLKE 358
 299 MLKPALARGELRCVGTATLDEYRQYTEKDAALERRPQVYVAEPSVDPDTISILRGLKE 358
 359 EGHGVRIQDRALINAQOLSAVITGRHLPDKAIDLVDEACANVRVQDSQPEEDNLER 418
 359 ATHGVQVILDSALIAAKKSHRYITDRQPLDKAIDLIDEAASRIKMEIDSKPEELDKR 418
 419 KRMQLEIELHALEREKDKASKARLIEVRKELDDRLDKLOPLMTKYRKEKERIDEIRRLKQ 478
 419 RLTLQKMLEAVKDEDEAGSKAEVTHLEKQIAEVEKEYNDLEEVKMAEKTIVGEGTQAV 478
 479 KREELMPSLOEARRVDLARAADLYCAIOVESAJAQL-ETSSSENVMLENTVGPHEI 537
 479 ELDKARTAFKAQREGDLAEARLQYGVIPLOKQLEQDEVAEENBEPKLIIRTKVTENEI 538
 538 AEVSVRTGIPVTRLGONEXERLIGLADRLHKEVGVQONAVNAVSAIILSRAGLGRAQ 597
 539 AEVVSATGIPVAKMMQGEREKULHMEEFVHDRVVGQDEAVAVAVNAVRSRAGLSDPNE 598
 598 PTGSFLPLGPTGVGKTELAKALAEQLFDDENLLVRLDMSEYMEQHSVRLIGAPPGYVGH 657
 599 PSQSFLPLGPTGVGKTELAKALAEQLFDDENLLVRLDMSEYMEQHSVRLIGAPPGYVGH 658
 658 EGGQLTEAVRRPYPYVILFEVEKAHVAVNTLLQVLDGRITDQCGRTVDPNSVILM 717
 659 EGGVLTEAVRRKPYVVLFEVEKAHPDVFNTLLQVLDGRITDQCGRTVDPNSVILM 718
 718 TSNLGAELHLAGLTKVTMVARDCVMEVRKHPPELLNRLDEIVVEDPLSHDQLRKVA 777
 719 TSNLGSQVRELGEAGTDDV-RTVNNVQSQHRPFIINRDELVIHSLKKAQIRGTA 777
 778 RLQMKDVAVLAERGVAVLATAADLYLAESVDPPYVGARPIRRWMEKVVVTELSKMWVR 837
 778 DIQLDRLSRLVDRLMSLTVDSDSAFLLIDAGFPDPPYVGARPLKRAIQQVNTLAQKILS 837
 838 BEIDENSTVIYIDAGDLVY 857
 838 GDFVAGDTILVKGNGHLVF 857

Patent No. 6355411
 GENERAL INFORMATION:
 APPLICANT: Ausubel, Frederick
 APPLICANT: Goodman, Howard M.
 APPLICANT: Rahme, Laurence G.
 APPLICANT: Mahajan-Miklos, Shalina
 APPLICANT: Tan, Man-Wah
 APPLICANT: Cao, Hui
 APPLICANT: Drenkard, Eliana
 APPLICANT: Tsongalis, John
 TITLE OF INVENTION: VIRULANCE-ASSOCIATED NUCLEIC ACID
 TITLE OF INVENTION: SEQUENCES AND USES THEREOF
 FILE REFERENCE: 00786/361002
 CURRENT FILING DATE: 1998-11-25
 PRIOR FILING DATE: 1997-11-25
 NUMBER OF SEQ ID NOS: 437
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 281
 LENGTH: 859
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-199-637A-281

Query Match 46.9%; Score 2133; DB 4; Length 859;
 Best Local Similarity 49.5%; Pred. No. 1e-169;
 Matches 427; Conservative 176; Mismatches 245; Indels 14; Gaps 6;
 QY 1 MNPEKTHKNTIATATHELANVAGHAQFTPLHLAGALISDPTGIFPOAISSAGGENAAQ 60
 DB 6 MRIDLTSKLQALSDAOLAMGKHDTATAGIHILSTLLEPSNI--SLLOQAGAR--LP 65
 QY 61 SAERVINOALKPLSOPSPDDIPASSSLIKVIRRAQAQKSRGDTLHVLQIMGLLE 119
 DB 66 RSG--LNKELDAPKIQSPGTGVNLSQDIARLLNQADRLAQKGDQFISSELVLAAMDE 123
 QY 120 SQIRDLNNEVGATARKVSEKLGKGGKVESAGDTNFQALNTYGRDLVQEA--GK 177
 DB 124 NTRLGKLLGGVSRKALANAVANLRG--GEAVNDPNVEESRQALDKYTVDMTKRAEKG 181
 QY 178 LDPVIGRDEIRVRILSRRTKKNPVLLIGEPGVGTAVVEGLAQRIVKGVDPNSLTDV 237
 DB 182 LDPVIGRDEIRRTIQVLSRRTKKNPVLLIGEPGVGTAVVEGLAQRIVKGVDPNSLTDV 241
 QY 238 LISDMGALVAGAKYRGEFEERLKSVLKEVEDAEAGKVLFIIDELHVLGAGKTEGSDAA 297
 DB 242 LLALDMGALVAGAKYRGEFEERLKSVLKEVEDAEAGKVLFIIDELHVLGAGKTEGSDAA 301
 QY 298 NLKPMPLARGOLRCIGATTILEYRKYVEKDAAFERRPQQVYVAEPSVDPDTISILRGLKE 357
 DB 302 NMLKPALARGELRCVGTATLDEYRQYTEKDAALERRPQVYVAEPSVDPDTISILRGLKE 361
 QY 358 YEGHGVRIQDRALINAQOLSAVITGRHLPDKAIDLVDEACANVRVQDSQPEEDNLE 417
 DB 362 YEVHGVSVITGALIAAKKSHRYITDRQPLDKAIDLIDEAASRIKMEIDSKPEELDKR 421
 QY 418 KRMQLEIELHALEREKDKASKARLIEVRKELDDRLDKLOPLMTKYRKEKERIDEIRRLK 477
 DB 422 RELIQLKIERALKKEDDEATRKELAKLEEDIVKLEREYADLEIEWSEKAQVQSGAQIQ 481
 QY 478 KQRELMPSLOEARRVDLARAADLYCAIOVESAJAQL-ETSSSENVMLENTVGPHEI 537
 DB 482 QKIEQAKQEMEAARKGDLSEMARIQVOTIPDLERSLQMDVMDHOKTENQLLNKVTDEEI 541
 QY 538 AEVSVRTGIPVTRLGONEXERLIGLADRLHKEVGVQONAVNAVSAIILSRAGLGRAQ 597
 DB 542 AEVSVRTGIPVTRLGONEXERLIGLADRLHKEVGVQONAVNAVSAIILSRAGLGRAQ 601
 QY 598 PTGSFLPLGPTGVGKTELAKALAEQLFDDENLLVRLDMSEYMEQHSVRLIGAPPGYVGH 657
 DB 602 PSQSFLPLGPTGVGKTELAKALAEQLFDDENLLVRLDMSEYMEQHSVRLIGAPPGYVGH 661


```

422 RRLIQIKTERALKKEDDEATKRKLAKLEEDIVKLEREVADLEDIWKSEKAQVQSSAQIQ 481
478 QKREBLMFSLQEAERYDLARAADLRYGAIQIEVESIAIQLEGTSSSENVMLTENVGPEHI 537
482 QKIEQAKQEMEAAREKGDLESMAIRIQYQTIPOLESLQWVDQHGKTENQLLRNKVTDDEI 541
538 AEVVSRTWICPIVTRLQNEKERLIGLADLRHKRVVQNOAVNAVSAILRSRAGLGRAQQ 597
542 AEVBSKWTCIPVSKMLGREGREKLIRMEQELHRVIGQDEAAVAVNAVRSRAGLADPNR 601
598 PTGSEFLFLGPTGVGKTELAKALAEQLEDDENLLVRLDMSEYMEQHSVSRLIGAPPGVVG 657
602 PGSGFLFLGPTGVGKTELCKALAEFLFDTEALVRLDMSEFMEKHSVARLIGAPPGVVG 661
658 EGGQQLTRAVRRPVCVILFDEVEKAHVAFNTLLQVLDGRLTDCGRVDFRNSVLIIM 717
662 EGGGYLTEAIRRKPYSVLLDEVEKAHPDVFNILLQVLEDGRLTDSHGRVDFRNTVVVM 721
718 TSNLGAHL--LAGLTGKVTMEVARDVMREYKHPERPELLNELDELTVVDFPLSHDQLRK 775
722 TSNLSAQIQEELAG-----DREAQRAAVMDAVNAHPRPEFINRIDEVVVEPLAREQIAG 776
776 VARLQMKQVAVRLAERGVALAVTDAALDYILAESYDPVYGAREIRRMMEKKVVTLSKMV 835
777 IAEIQGLRLKRLAEREISLELSQEAELDKLIAVGDFDVGARFLKRAIQORWIENPLAQLI 836
836 VREEIDENSTVIDAGADLVY 857
837 LAGFAPGASISAKVEGDEIVF 858

RESULT 7
US-09-134-001C-3979
; Sequence 3979, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3979
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3979

Query Match 46.3%; Score 2108; DB 4; Length 871;
Best Local Similarity 53.4%; Pred. No. 1.3e-167;
Matches 415; Conservative 167; Mismatches 173; Indels 22; Gaps 8;

QY 86 SSSLIKVIRRAQAQKSRGDTHAVDOLINGLEDQSQIRDLINVEGVATARKSEVEKLR 145
DB 88 SAKTNELLDKAEKYMSEYDEFISMEHILRAADTDETTQ--KWVGNKVEVIKEIITKVR 145
QY 146 GKGEKKVESAGDTNFQALKTYCRDLEVEQA--GKLDVIGRDEIRVVRILSRRTKNKP 203
DB 146 G--GNHVTQNPENVTEALKEKYGRDLVVEVQKQNDPVIQRDEIRNTILSRKTKNP 203
QY 204 VLITGEPGVKTAVVEGLAQIRVIRKGVDPNSLTDLRLISLDMGALVAGAKYRGEFEERLKS 263
DB 204 VLITGEPGVKTAIVEGLAQIRVIRKVDVPESLDDKTIPEDLSALVAGAKYRGEFEERLKAV 263
QY 264 LKEVEDABKVLIFIDEIHLVLGAGKTEGSDAANLFKPMRLARGQLFCIGATTILEEVRKY 323
DB 264 LKEVKESGRILFIDEIHLVLGAGKTDGAMDAGNMLKPMRLARGELFCIGATTILNEYREY 323

```

324 VEKDAAFERRFOQVYVABSPVDTISILRLGLKEKYEHHGVRIODRALINAAQASRYIT 383
 324 IKDSALERRFOKVGSPDVENTISILRLGLKERVYVGVRIODRALVAAASDRYIT 383
 384 GHLPDKAIDLVDKACANRVOLDSPQBEIDNLERKRMQLEIELHALEREKDKASKALI 443
 384 DRFLPDKAIDLVDQACATIRTEMGNSPTDELQVNRVRVQLEIEESALKNESDNASKRLE 443
 444 EYRKELDLARLQPLTKYRKERIDEIRLKKREELMPS---LQEAERRVDLARA 500
 444 ELQELNEKEKQSSLSKSVQEKK---IAVQEKRAELSSQALDEDAQTEGNLEKAA 500
 501 DLRYGAIQVESAIQLE-----GTSSEENVMLTENVEPHIAEVVSWTGPVTRLQON 555
 501 ELQYGTIPQLEKELQEFQEFQEDGETSEDESMIREVVSDEIGDIVSQWGTGIPVSKLVET 560
 556 EKERLIGLADRLHVRVQONAVNAVSEAILSRAGLGRAOQTSPLFGTGVGKTEL 615
 561 EREKLISLSDILHVRVQDKAVDLVSDAVVRARAGIKDPNRPISFLFGTGVGKTEL 620
 616 AKALABQLEDDENLVRIDMSYMEQHSVSRILGAPPGVVGHEGGQLEAVRRPYPYCVI 675
 621 AKSLASLSDSKHMRIDMSYMEKHAVSRILGAPPGVVGHEGGQLEAVRRPYPYCVI 680
 676 LFDEVEKAHVAVFNTLLQVLDGRLTDGQRTVDPRNSVIIMTSLNLAGAHLGAGTKYT 735
 681 LLDEVEKAHSDVFNLLQVLDGRLTDGQRTVDPRNSVIIMTSLNLAGAHLGAGTKYT 738
 736 MEVARD---CVREVKHFRPELLNRLDEIVFDELPHDQKVARLQMDVAVRLAERG 792
 739 GEISDTEKAVMSDLHAYKPEILNMDDIVLFKPLSVDDMSMIVDKILTQNLMLLDQH 798
 793 VALAVTDAALDYILABSPVVGARPIRMEKVKVTELSKMWVREEDENSTVVID 849
 799 ISIEVTEAKKWLGEAYEPQFARPLKRFVQRIETPIARMWIKESLPEGTLIKVD 855

us-09-198-452A-163
 Sequence 163, Application US/09198452A
 Patent No. 6559294
 GENERAL INFORMATION:
 APPLICANT: Griffiths, R.
 TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention of infection and treatment of infection
 FILE REFERENCE: 9710-003-999
 CURRENT APPLICATION NUMBER: US/09/198,452A
 CURRENT FILING DATE: 1998-11-24
 NUMBER OF SEQ ID NOS: 6849
 SEQ ID NO 163
 LENGTH: 872
 TYPE: PRT
 ORGANISM: Chlamydia pneumoniae
 -09-198-452A-163

Query Match 42.0%; Score 1912; DB 4; Length 872;
 Best Local Similarity 46.5%; Pred. No. 3.2e-151;
 Matches 404; Conservative 171; Mismatches 271; Indels 22; Gaps 8;

4 EKPTHKNTETATHELAVNAGHAQFPLHAGALISDPTGIFPQAISSAGGNAQAQSAE 63
 8 EKFSDAVSEALEKAFELAKSGKSYTVVTHLLALLENTESFLYVLIKDIHGNFG----- 62
 64 RVINOALKKLQSPQSP---PDDIPASSSLIKVIRRAQAQKSGDTHLAVDQIMGLLE 119
 63 -LLNTAVKDALSRFTVVEGVDPKPSGGLQTLRDAKQAKTLGDEYISGDHLLAFWS 121
 120 DSQ---IRDLLNEVGAVTARVKEVEKELKGGKKEVSGASGDTNPAQTKYGRDLVQA-- 175
 122 SNKEPFSWKTQTVKSPKDLKLNLTIR--RGNRMDSPSAESNFOGLEKYCKNLTALARE 179
 176 GKLDPVIGRDEIRRVILSRRTKNNPVLIGERGCVKTAVEGLAQRIKVGDPNSLTD 235

180 GKLDPVIGRDEIRRTIQVLSRRTKNPMLEPGVGKTAIEGLAIRLIQGVDPSELKG 239
 236 VRLISLDMGALVAGYRGFEERLKSVLKEVEDABKVLFTDEIHLVLGAGKTEGSM 295
 240 KQIYVLDMGALVAGYRGFEERLKSVLKVESGGEHIFIDEVHTLVGAGATGAMD 299
 296 AANLKPMLARGQLRCIGATTLEBYKXYVEKDAAFERRFOQVYVABSPVDTISILRLGLK 355
 300 AANLLPALARGTLHCIGATTLEBYKXYVEKDAAFERRFOQVYVABSPVDTISILRLGLR 359
 356 EKYEGHGVRIODRALINAAQASRYITGRHLDPDKAIDLVDKACANRVOLDSPQBEIDN 415
 360 EKYEIFGVRIETGALNAAVLLSYRIPDRFLPDKAIDLDEAASLIRWQIGSLPLPIDE 419
 416 LERKRMQLEIELHALEREKDKASKARLIEVRKELDDLRLQPLTKYRKERIDEIRR 475
 420 KERELAALIVQKQEAIRKQSPSYQEEADAMQKSIDALRELASRLGWDEBKLIQGLKE 479
 476 LKQREELMPSLQEAERRYDLARAADRLRYCAIQEVESATQLEGT--SSEENVMLTENVGP 534
 480 KONSLESMTSEBEAERVADYNRAELRYSLIPQLEBEIKQDEASLNQRDLRLQBEVDE 539
 535 EHTAAYVSRVTGIPVTELGQNEKERLIGLADRLHVRVQONAVNAVSEAILSRAGLGR 594
 540 RLTAQVNVANWTGIPVQKLEGEAKLILIEESLEERVVGQPPFAVSVDISIRARVGLND 599
 595 AQQPTGSLFLGPTGVGKTELAKALAFQLEDDENLVRIDMSYMEQHSVSRILGAPPGY 654
 600 PQRPLGVFLFGPTGVGKTELAKALADLLFNKEAMVYRPMSEYMEKHISIKLIGSPGY 659
 655 VGHEGGQLEAVRRPYPYCVILFDEVEKAHVAVFNTLLQVLDGRLTDGQRTVDPRNSV 714
 660 VGYEGGSLSEALRRRPSYVLFDETEKADKEVINILLQVFDGGLTDGKKRKNCKNAL 719
 715 IIMTSLNLAGAHLGAGTKYVMEVARDVNRV---RKHFRPELLNRLDEIVVDFPLSH 770
 720 FIMTSLNIGSPE--LADYCKSGSELTKEAILSVSPVLKRYLSPEFFMNRIDEILPFVPLTK 778
 771 DOLRKVARLQMDVAVRLAERGVALAVTDAALDYILABSPVVGARPIRMEKVKVTE 830
 779 EDIVKIVGICQMRIRIAQRLKARRINLSWDDSVILFLSQGYDSAFGARPLKRLIQCKVIL 838
 831 LSKMWVREEDENSTVVIDAGAGLVYR 858
 839 LSKALLKGIDKPDTSIELTWAKEVLVFK 866

RESULT 9
 US-08-249-380-2
 Sequence 2, Application US/08249380
 Patent No. 5827685
 GENERAL INFORMATION:
 APPLICANT: Lindquist, Susan
 TITLE OF INVENTION: Methods and Compositions of Genetic
 TITLE OF INVENTION: Stress Response Systems
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P. O. Box 4433
 CITY: Houston
 STATE: TX
 COUNTRY: USA
 ZIP: 77210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/249,380
 FILING DATE:
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/710,187
 FILING DATE: 31-MAY-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Parker, David L.
 REGISTRATION NUMBER: 32,165
 REFERENCE/DOCKET NUMBER: ARCD:024
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 713-787-1400
 TELEFAX: 713-789-2679

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:
 LENGTH: 908 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

8-249-380-2

Query Match 40.4%; Score 1839; DB 2; Length 908;
 Best Local Similarity 42.2%; Pred. No. 4.4e-145; Indels 80; Gaps 18;
 Matches 399; Conservative 181; Mismatches 286;

1 MNPE-KPTHKNTETATIAHELAVNAGHAQTPHLAALISDP-TGIFPOAISSAGENA 58

1 MNDQTFERALTITLTAQKLASHDHPQLQPIHILAAFIETPDGSPVYL-----QNL 54

59 AQSA-----ERVINQAKLPSQSPDDIPASSSLIKVIRPAQAQKSRGDTHLAVD 111

55 IEKGRYDYLFPKVVNRNLVRIPQQQAPAEITPSYALGVQLQDAKIQKQKDSFIAD 114

112 QLIMGLLEDSQIRDLNNEVGAVATARKVSEVKLRGKKGKVESASGDTN--FOALKTYGR 169

115 HILFALFNDSIQIFKEAQVDIEAIKQALELGR--NTRIDSRGADTNPLEVLSKYAI 172

170 DLVEQA--GKLDPVIGRDEETIRRVVIRLSRRTKKNPVLIGEPGVGKTAVVEGLAQRIKVG 227

173 DMTQARQGLDPVIGREETIRVIRLARIKSNPCILGEPGIGKTAITEGVAQRIIDD 232

228 DVPSNLTDVRLISLDMGALVAGKVGFEERLRSKLVKEVEDAESGKVLITDETHLVGA 287

233 DVPTILOQAKLFLSLDALTAAGAKYKGDFFERFRGVLKEIEESKTLVIFIDEIHLUMGN 292

288 KTEGSDMAANLFPMLARGQLRCIGATTLEEYKRYVEKDAAFERRFPQVYVAPSPVDT 347

293 GKD---DAANILKPALSRGQLKVI GATTNNEYRSIVKOGAFERRFKIEVAEPSVRQT 348

348 ISILRLKEKVEGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLD 407

349 VAILRGLQPKYIEHGVRIILDSALVTAQAALAKRYLPYRRLPDSALDLVDISCAGVARD 408

408 SQPEIDNLERKRMQLEIHLALEREKDKASKARI---IEVRKELD-----DLRDLKQ 457

409 SKPELDSKEQS-----TAIDSRDKSRREVERCKLTKRLKLARQKVASLQEELE 460

458 PLTWKYRKEKERIDEIRLKRQBELMESLOEABRRYDLARAADLRYCAIQVESALIAQL 517

461 PLRQYNEEKHGHEELTQAKKLDLENKALVAERYDTRTAADLRYFAIPDIKKQIEKL 520

518 EGTSESE-----NVMLTENVGPEHIAEVRSWTGIPVTRIGQNEKRLGLADRLHKRV 571

521 EDQVAEERRRAGANSMIQNVVDSDTTISETAARLTGIPVKLSSENEKLIHMERDLSSEV 580

572 VQONQAVNASAILRSRAGLGRAOQPTGSLFGLPTGVGKTELAKALAEOLFDDENLLV 631

581 VQMDAIKAVSNVRLSRGLANRPQ--ASFLLGLSGSGKTELAKKAVAGFLFNEDNMI 639

632 RIDMGEYMEQHSVRLIGAPPYGVGHEGGQLTEAVRRRPYCVILFDEVEKAHVAVNTL 691

640 RYDCSELSEKAVSKLLGTAGYGVYDEGGFLTNQLQYKPYSVLLFDEVEKAHPDVLTV 699

692 LQVLDDGRLTDQGRVTDFRNSVIITMNLGAELLAGLTKVMTWEVARDCCVMREVRKHF 751

Db 700 LQMLDDGRITSGQGTIDCSNCIVIMTSLNLAELFINSQQSGSKI-QESTKNLVMGAVRQHF 758
 QY 752 REELNRLDEIVFDPDLSDOLRKVARLQMKDVAVRLA--ERGVALAVTDAALDYLAES 809
 Db 759 RPEFLNRISIVIFNKLRSKAIHKIVDLKEIEERPFQNDKHYKLNLTQEAOKDFLAKVG 818
 QY 810 YDPVYGARPIRMMEKKVVTLSKMWREEDIDENSTVYDAGAGDLVTVRVESGGVLVDAST 869
 Db 819 YSDDMGARPLNRILQNEILNKLALRILKNEIKDKETVN-----VVLKK 861
 QY 870 GKSDVLTHIANGPKESDAQAV-----KKMREIEDDDNEEMIE 910
 Db 862 GKSRD-----ENVPEAECEULEVLPNHEATIGADTLGGDDNEDSME 902

RESULT 10

US-09-134-001C-4081

; Sequence 4081, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 4081

; LENGTH: 823

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4081

Query Match 40.2%; Score 1830; DB 4; Length 823;
 Best Local Similarity 44.0%; Pred. No. 2.2e-144;
 Matches 386; Conservative 156; Mismatches 236; Indels 100; Gaps 13;

QY 5 KPTHKNTETATIAHELAVNAGHAQTPHLAALISDPGIFPOAISSAG--GENAQA 62

Db 11 RLTERAQRVLAAQEAELALNHSNTGTEHLGLMKEPEGTAARKVLESFNTEDKVIEV 70

QY 63 ERVINQAKLPSQSPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLIMGLLEDSQ 122

Db 71 EKLIGHGQOMGTILHYP-----RAKVIELSDWEARKLHHNFVGTIELLGLIRENE 123

QY 123 IRDLLNEVGAVT-----ARVSEVKEKLRG--KGGKKVESASGDTNFQALKTYGRD 170

Db 124 -----GVAARVFAFLDLNITKARQVVKALGSPMSNKAQAQKSNNTPTLDGLARD 175

QY 171 L--VEQAGKLDVIGDEEIRRVVIRLSRRTKKNPVLIGEPGVGKTAVVEGLAQRIVKGD 228

Db 176 LTVIAKDGTLDPVVGGRDKETRVIEVLSRRTKKNPVLIGEPGVGKTAAEGLAQAIVKNE 235

QY 229 VPNSLTDVRLISLDMGALVAGAKYGEFEERLRSKLVKEVEDAESGKVLITDETHLVGAG 288

Db 236 VETLKDVRVSLDMGTVVAGTKYRGEFEERLKKVMEELHQA--GNVILFIDELHVLVAG 294

QY 289 KTEGSDMAANLFPMLARGQLRCIGATTLEEYKRYVEKDAAFERRFPQVYVAPSPVDTI 348

Db 295 GAEGATDASNLKALPARGELQICGATTILDEYRKNIEKDAALEREFPQIQVDEPTVEBTI 354

QY 349 SILRLKEKVEGHGVRIQDRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQLDS 408

Db 355 EILKGLRDRYEAHRRINISDEALEAAKLSRYSDRFLPDKAIDLIDEASKVRLKSH 414

QY 409 QPEIDNLERKRMQLEIHLALEREKDKASKARLIEVRKELDDLRDLKQLPTMKYRKEKE 468

Db 415 TFSNLKEIEQ-----EIDKVKNEKDAVHAQEFENAAANLRDKQSKL----- 455

825 KQ 826

ULT 12

09-621-855-2

Sequence 2, Application US/09621855

Patent No. 6346608

GENERAL INFORMATION:

APPLICANT: Jaworski, Deborah J.

Wang, Min

Shilling, Lisa K.

Burnham, Martin

Fosberry, Andrew

Hodgeson, John E.

Lawlor, Elizabeth

Rosenberg, Martin

Ward, Judith

TITLE OF INVENTION: MecB

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESS: Dechert, Price & Rhoads

STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103-2793

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/621,855

FILING DATE: 24-Jul-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/040,843

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Falk, Stephen T

REGISTRATION NUMBER: 36,795

REFERENCE/DOCKET NUMBER: GM10082

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-994-2488

TELEFAX: 215-994-2222

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 866 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

09-621-855-2

Query Match

Best Local Similarity 39.8%; Score 1810.5; DB 4; Length 866;

Matches 385; Conservative 164; Mismatches 252; Indels 101; Gaps 14;

5 KETHKTETIATAHLAVNAGHAQTPHLGALISDPTGIFPQAISAGGENAAQSAER 64

5 RUTERARVLAHQAEARLHNSNIGTEHLLGLMKPEPIAAKVL-----ESFNITSDK 59

65 VINOAKKLPSQPPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQIMGLLEDQSQR 124

60 VIEVEKLIGHQDVGTHLYTPRAKKVIELSMDERKLHNFVGTVHLLGLIRENE-- 117

125 DLLNEVGVAAT---ARVKSVEKLKQEGK-----KVESAGDTHNFQALKTVGRDL- 171

118 -----GVAARVFANLDLNTIKARARVVKALGNPEMSNKAQAQSKSNNTPTLDSLRDLT 171

172 -VEQAGKLDLPVIGREEIRRVVRLISRRRTKNNPVLIGEPGVGKTAVVEGLAQRIKVGDPV 230

Db 172 VIAKGTLPVIGRDKETITRVIEVLSRRRTKNNPVLIGEPGVGKTAIAEGLAQIAVNVVP 231

QY 231 NSLTDVRLISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEGKVLFIIDEIHLVLGAGKT 290

Db 232 ETLKDKRVMSLDNGTIVAGTKYRGEFEERLKKVWEIQQA-GNVILFIDELHTIVGAGA 290

QY 291 EGSMDAANLKPMLARGQLRCIGATTLEIYRYKVEKDAAPERRFQQVYVAEVPDITSI 350

Db 291 EGAIIDASNTLKPALARGELQCIGATTLDEYRKNIKDAALERRFPQVQVDEPSVVDTVAI 350

QY 351 LRGLKEKYEGHGVRIODRALINAAQLSARYITGHLDPKADLDVDEACANVRVQLDSOP 410

Db 351 LKGLDRKYEAHRINISDEAIEAAVKLSNRYSDRFLPKAIDLIDEASSKVKLRKSHHTP 410

QY 411 EEIDNLERKMQLEIEIHALEREKOKASKARLIEVRKLEDDLDLKLQPLTKMYRKEKERI 470

Db 411 NNLKEIEQ-----EIEKVKEKDAAVHAQFE---NNAANLRDKQTLEKQY----- 453

QY 471 DEIRRLKQKREELMFSLOABRRYDLARAADLRYGAIQBVESAIAOLEGTSSSEENVMLTE 530

Db 454 -----EEAKNEW-----KNAQNGMST- 469

QY 531 NVGPEHLAEVSWHTGIPVTRLGQNEKERLIGLADRLHKRVVGQCAVNAVSEAILRSRA 590

Db 470 SLSEEDIAEVIAGTGIPLTKINETESEKLLSLEDTLHERVIGQKQDANVSIKAVRARARA 529

QY 591 GLGAAQQTGSGFLGPTGVGKTELAKALAEQLFDDENLLVRIDMSEYMEQHSVSRLLIGA 650

Db 530 GLKDPKRPISGFIPLGPTGVGKTELARALAESNMGDDAMIRVDMSEFMEKHAVERLVGA 589

QY 651 PPGYVCHERGQQLTEAVRRPYCVILFDEVEKAHVAVFNTLLQVLDGRLTGQGTVDVF 710

Db 590 PPGYVGHDDGQQLTEKVRKPKSVILFDEIEKAHPDVFNILLQVLDGRLTGQGTVDVF 649

QY 711 RNSVLIWTSNLGAHL---LAGLTCK---VTMEVARDCVMEVRKHPRPELLNRLDEIV 763

Db 650 RTIILIIISNVGAQELQDQRFAGFGSSDGDYETITKMLKELKNSFRPEFLNRVDDII 709

QY 764 VFDPLSHDQLRKVARLQMKDVAVRLAERGVALVATDAALDYTLAESYDPVYGARPIRRM 823

Db 710 VFHKLTKELKEIVTMVWVKLTNRLSEQNINIIVTDKAKDKIAEGYDPEYGARPLIRAI 769

QY 824 EKKVTVELSKAVVREEDENSIVYIDAGAGDLVYRVESGLVDASTGKKSVDVLIHANGP 883

Db 770 QKTIEDNLSELIDGNQIEGKKTVDVDHGKFKY-----DIAEOTSETKTPSQAXILXNSP 824

QY 884 KR 885

Db 825 KQ 826

RESULT 13

US-09-252-991A-22493

; Sequence 22493, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 22493

; LENGTH: 1044

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-22493

Query Match 39.3%; Score 1789.5; DB 4; Length 1044;
Best Local Similarity 43.8%; Pred. No. 7.6e-141; Indels 89; Gaps 19;
Matches 397; Conservative 168; Mismatches 252;

4 EKFTHTNFTIATAHELAVNAGAOFTPLHLGALISDPTGIFPOAISAGGENAQAQSAE 63
199 EYFSKOTTEYLQRAQVAEAFEGKREVDTEHLLYAL-ADADV--QAVLKQFGLSPAELKQ 255
64 RVINQAKKLPGQPPDDIPASSLSLKVIRRAQAQKRGDTHLAVDOLMGL--LEDS 121
256 YIEANAVRG-ASKGEASEDMTISPRVKSALQAFALSRELGSYVGPHEHLLGLAAVPS 314
122 QIRDLINAVGATARKVSVEKLRGKGGKVESASGDTNFQALKTVGRDLVQEA--GKLD 179
315 FAGTLKKYGLTEQALQKAVKVVGK-GAEDRGVDPSTPQDKFSRDLTLAREGKLD 373
180 PVIGRDEIRRVVRLSRRTKNPNVLIGPGVGKTAIVEGLAQRIKVGDPVNSLTDVRLI 239
374 PVIGRSKEVETTIEVLARKKNPNVLIGPGVGKTAIVEGLAQRIKVGDPVNSLTDVRLI 433
240 SLDMGALVAGAKRGFEERLKSVLKEVEDAGKVLFIIDEHLVLGAK--TEGSMQAA 297
434 ELNINAVAGAKRGFEERLKSVLKEVEDAGKVLFIIDEHLVLGAK--TEGSMQAA 493
298 NLFKPMIARGQLRCIGATTLEEVKRYKVEKDAFERFQVYVAEPVPTISILRLKKEK 357
494 NVLKPAMARGEMNLICATTLNEVQKIEKDALEFRFPVFPVETVEGTISILRLKKEK 553
358 YEGHGVRIQDRALINAAQLSARYITGRHLPKAIDLVDCAANVRVQLDSQPEIDNLE 417
554 LEGHHKVTIRDEAFVAAEELSDRYIGNRFLPKAIDLIDQAAARVIASTSPAEI--- 609
418 RKEMQLEIHLALEREKDKAKARLIEVREKELDLKQLPLTMKYRKEKERIDIRLK 477
610 ---QLEAEALQKREQDYAAGKWKYDEAKVFE-----KKIQ 643
478 QKREELMFSIQEAERRYDLARAADLYGAIQEVESAIAQLEGTSEENVMTENVGPEHI 537
644 ERKEHL-----EQITERWQQTQSKTEE-----VRVEDI 672
538 AEVVSMTGIPVTRLGONKERLIGLADRLHKKRVGQNOAVNAVSEAILRSAGIQAQQ 597
673 AETISRLTGIPVTELTAEERELQMEERLHORVIGQEAITAVSDAVRLARAGLRQGR 732
598 PTGSFLFGPTGKTELAKALAEQIFDENLIVRDMSEYMEQHSVSELICAPPQYVCH 657
733 PIATFLFGPTGKTELAKALAEVVGDEDMIRIDMSEYMERHVAVSRLIAGPGYVGY 792
658 EGGQLTEAVRRPYPYVILFDEVEKAVAVFNTLLQVLDGRLTDGQRTVDPRNSVIIM 717
793 DEGGQLTEAVRRPYPYVILLDEIEKAHADVNILLQVFDGRLTDGGRVVDFTNIIIA 852
718 TSNLGAHLL-----AGLTGKVTMEVARDVCMREVRGHPPELLNRLDEIVDPDLSHDL 773
853 TSNLGSSELIMKNAQAQGEFAQPEKLRK-LMTTLRGHFRPEFLNRLDEIVPESLSKAQI 911
774 RKVARKQMDVAVRLAERGVALAVTDAALYILAESVDVYVGARPTRRMWEKKVVTLSK 833
912 EDVRLQLEKVKGAHAHQDILYHIDSLVGLHAEAYQPEFGARELKRQIRQOETRLAT 971
834 MVVREIDENSTV--YIDAGAGLVVRVESGGIADASTGKSPVLTHIANGPK-RSDAAQ 890
972 AMLKGEVKEGETVTFYDADKG-VGYR--KGAAPKPAARKKSGA---GETPKGRATAAR 1024
891 --AVKK 894
1025 KPAACK 1030

USULT 14
-09-040-843-4
Sequence 4, Application US/09040843
Patent No. 6124119

GENERAL INFORMATION:
APPLICANT: Jaworski, Deborah J.
APPLICANT: Wang, Min
APPLICANT: Shilling, Lisa K.
APPLICANT: Burnham, Martin
APPLICANT: Fosberry, Andrew
APPLICANT: Hodgson, John E.
APPLICANT: Lawlor, Elizabeth
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
TITLE OF INVENTION: MecB
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,843
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/057,535
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM10082
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 672 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-040-843-4

Query Match 38.4%; Score 1745.5; DB 3; Length 672;
Best Local Similarity 49.0%; Pred. No. 1.9e-137;
Matches 352; Conservative 129; Mismatches 166; Indels 71; Gaps 9;

QY 151 KVESASGDTNFQALKTVGRDL--VEQAGKLDPVIGRDEEIRRVRLSRRTKNPNVLIGE 208
DB 4 KNAQASKNNPTLDSLRDLTVAKDGTLDPVIGRDEEIRRVRLSRRTKNPNVLIGE 63
QY 209 PGVGTAVVEGLAQRIYKVGDPVNSLTDVRLISLDMGALVAGAKYRGFEERLKSVLKEVE 268
DB 64 PGVGTAAEGLAQIAVNVNEVPELTKDKRVMSLDMGTTVAGTKYRGFEERLKSVMEEIQ 123
QY 269 DAEKVLFIIDEHLVIGAGKTEGSDMAANLFPKMLARGQLRCIGATTILEEYKYVEKDA 328
DB 124 QA-GNVILFIDELTLVAGAGAGCAIDASNILKLPALARGELCIGATTLLDEYRKNIEKDA 182
QY 329 AFERRFOQVYVABSPVPTTISILRLKKEKVEGHHGVRIQDRALINAAQLSARYITGRHLP 388
DB 183 ALERRFQVQVDEFSVVDVTVAILLKGLDRDYEAHHRINISDEATEAAVKLSNRYVSDFLP 242
QY 389 DKAILDVEACANVRVQLDSQPEIDNLERKRMQLTELHALEREKDKASKARLIEVRKE 448
DB 243 DKAILDVEASSKVRKLSHTTFFNNLKEIEQ-----EIEKVKNEKDAAVHAQBEF---N 292
QY 449 LDDLRDKLQPLTMKYRKEKERIDEIRLKKREELMFSIQEAERRYDLARAADLYGAIQ 508


```

293 AANLRDKQTKLEKQY-----EAKNEW----- 314
509 EVESATIAQLGEGTSSENNVMTENGVGPEHIAEVSVRWTGIPVTRLGQNEKERLIGLADRLH 568
315 -----KNAQNGMST-SLSEEDIAEVIAGTGIPLTKINETESEKLSLEDTLH 361
569 KRVVGQNAVAVSEAILRSRAGLGRAGQPTGSFLPLGPTGVGKTELAKALAEQLEDDEN 628
362 ERVIGQKDAVNSISKAVRRARAGLKDPRIGSFIFLPGTGVGKTELARALAESMFGDD 421
629 LLVRIDMSYMEQHSVRLIGAPPVGVGHEGGQLETEAVRRRPYCVILFDEVEKAHVAVF 688
422 AMIRVDMSEFMKHAHVSRLVAPPVGVGHDDGQLETKVRRKPYSVILFDETEKAHPDV 481
689 NTLLQVLDGRLTDGGRVDFRNSVIMTSLGAHL-----LAGLTKG---VTMEVARD 741
482 NILLQVLDGHLTDTKGRVDFRNIIMTNSVGAQELQDQRFAGFGSSDQDYETIRK 541
742 CVMREVKHFRPELNLRLDEIVVDFPLSHDQLRKVARLQMDVAVRLAERGVALAVTDA 801
542 TMLKELKNSFRPELNLRLDIIIVFHKLTKEELKEIVTMVNVKLITNRLSEQNINIIVTDA 601
802 LDYILAESYDPVYGARPIRWMKVKVTVLSKXVREIDENSTVIIDAGAGDLVYR 859
602 KDKIAEGYDPEYGARPLIRAIQKTIEDNLSELIDGNQIEGKKVTVVDHDKGKFKYDI 659

```

09-621-855-4
 Sequence 4, Application US/09621855
 Patent No. 6346608
 GENERAL INFORMATION:
 APPLICANT: Jaworski, Deborah J.
 Shilling, Lisa K.
 Burnham, Martin
 Fossberry, Andrew
 Hodgson, John E.
 Lawlor, Elizabeth
 Rosenberg, Martin
 Ward, Judith
 TITLE OF INVENTION: Mech
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESS: Dechert, Price & Rhoads
 STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103-2793
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows 95
 SOFTWARE: FastSeq for Windows Version 2.0b
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/621,855
 FILING DATE: 24-Jul-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/040,843
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Falk, Stephen T.
 REGISTRATION NUMBER: 36,795
 REFERENCE/DOCKET NUMBER: GM10082
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-994-2488
 TELEFAX: 215-994-2222
 TELEX: <Unknown>
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 672 amino acids

```

;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-621-855-4

Query Match
Best Local Similarity 38.4%; Score 1745.5; DB 4; Length 672;
Matches 352; Conservative 129; Mismatches 166; Indels 71; Gaps 9;

QY 151 KVESASGDTNFQALKTKYGRDL--VEQAGKLDVPVIGRDEIRRVVRLISRRTKNNPVLIGE 208
Db 4 KNAQAKSNNTPTLSDSLARDLTVIADKDTLPVIGRDKETIRVIEVLSRRTKNNPVLIGE 63
QY 209 PGVCKTAVVGLAQRIYKGVDPNSLTIVRLISLDMGALVAGAKYRGEPFERIKSVLKEVE 268
Db 64 PGVCKTAVVGLAQRIYKGVDPNSLTIVRLISLDMGALVAGAKYRGEPFERIKSVLKEVE 123
QY 269 DAEGKVLFDIEHLVLGAGKTEGSDMAANLFKPLMARGQLRCIGATTLFEYRYKYVEKDA 328
Db 124 QA-GNVILFIDELHTLVGAGGAGDAIDASNILKPALARGELQICIGATTLDEYRKNIEKDA 182
QY 329 AFERRFOQVVASPSVDDTISILRLGKKEKTEGHGVRIODRALINAAQLSARYITGRHLP 388
Db 183 ALERRFOQVVASPSVDDTISILRLGKKEKTEGHGVRIODRALINAAQLSARYITGRHLP 242
QY 389 DKAIDLVDEACANVRVQDLSQPEIDNLERKRMQLELELHALEREKDKASKARLIEVRKE 448
Db 243 DKAIDLVDEACANVRVQDLSQPEIDNLERKRMQLELELHALEREKDKASKARLIEVRKE 292
QY 449 LDDLRDKLQPLTKMYRKEKERIDEIRRLKQRELMFSLQEAERYDLARAADRLRYGAIQ 508
Db 293 AANLRDKQTKLEKQY-----EAKNEW----- 314
QY 509 EVESATIAQLGEGTSSENNVMTENGVGPEHIAEVSVRWTGIPVTRLGQNEKERLIGLADRLH 568
Db 315 -----KNAQNGMST-SLSEEDIAEVIAGTGIPLTKINETESEKLSLEDTLH 361
QY 569 KRVVGQNAVAVSEAILRSRAGLGRAGQPTGSFLPLGPTGVGKTELAKALAEQLEDDEN 628
Db 362 ERVIGQKDAVNSISKAVRRARAGLKDPRIGSFIFLPGTGVGKTELARALAESMFGDD 421
QY 629 LLVRIDMSYMEQHSVRLIGAPPVGVGHEGGQLETEAVRRRPYCVILFDEVEKAHVAVF 688
Db 422 AMIRVDMSEFMKHAHVSRLVAPPVGVGHDDGQLETKVRRKPYSVILFDEIEKAHPDV 481
QY 689 NTLLQVLDGRLTDGGRVDFRNSVIMTSLGAHL-----LAGLTKG---VTMEVARD 741
Db 482 NILLQVLDGHLTDTKGRVDFRNIIMTNSVGAQELQDQRFAGFGSSDQDYETIRK 541
QY 742 CVMREVKHFRPELNLRLDEIVVDFPLSHDQLRKVARLQMDVAVRLAERGVALAVTDA 801
Db 542 TMLKELKNSFRPELNLRLDIIIVFHKLTKEELKEIVTMVNVKLITNRLSEQNINIIVTDA 601
QY 802 LDYILAESYDPVYGARPIRWMKVKVTVLSKXVREIDENSTVIIDAGAGDLVYR 859
Db 602 KDKIAEGYDPEYGARPLIRAIQKTIEDNLSELIDGNQIEGKKVTVVDHDKGKFKYDI 659

```

Search completed: February 13, 2004, 01:35:12
 Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Protein - protein search, using sw model

Search time 76 Seconds
(without alignments)
2509.832 Million cell updates/sec

US-09-812-350-17

Score: 4550

Sequence: 1 MNPEFTHKTNETATAEH.....VKMKRIEIEDDDNEEMIED 911

BLSUM62

Gapop 10.0 , Gapext 0.5

801455 seqs, 209382283 residues

1 number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Alt	Score	Query Match	Length	ID	Description
1	3883.5	85.4	912	12	US-10-310-154-481
2	2315	50.9	881	12	US-10-369-493-20047
3	2288	50.3	872	12	US-10-310-154-484
4	2288	50.3	872	12	US-10-369-493-2762
5	2288	50.3	872	14	US-10-047-260-38
6	2282	50.2	879	15	US-10-156-761-14770
7	2264.5	49.8	857	12	US-10-369-493-20488
8	2247	49.4	862	12	US-10-369-493-19762
9	2243	49.3	862	12	US-10-369-493-20953
10	2234.5	49.1	905	12	US-10-369-493-22555
11	2225	48.9	868	12	US-10-369-493-11982
12	2224	48.9	859	12	US-10-369-493-16834
13	2214.5	48.7	907	12	US-10-369-493-3990
14	2204	48.4	861	9	US-09-815-242-14088
15	2199.5	48.3	858	12	US-10-369-493-10608

Sequence 11393, A
Sequence 7831, Ap
Sequence 47, Appl
Sequence 483, App
Sequence 10258, A
Sequence 23508, A
Sequence 17826, A
Sequence 2769, Ap
Sequence 15599, A
Sequence 15976, A
Sequence 16333, A
Sequence 11134, A
Sequence 485, App
Sequence 17514, A
Sequence 8096, Ap
Sequence 9166, Ap
Sequence 9374, Ap
Sequence 6999, Ap
Sequence 5230, Ap
Sequence 12526, A
Sequence 19403, A
Sequence 19111, A
Sequence 5192, Ap
Sequence 19625, A
Sequence 281, App
Sequence 20217, A
Sequence 12048, A
Sequence 558, App
Sequence 13666, A
Sequence 18448, A

ALIGNMENTS

RESULT 1

US-10-310-154-481
Sequence 481, Application US/10310154
Publication No. US20030233670A1

GENERAL INFORMATION:
APPLICANT: Edgerton, Michael D

APPLICANT: Chomet, Paul S.

APPLICANT: Adams, Thomas H

APPLICANT: Ruff, Thomas G.

APPLICANT: Agarwal, Ameeta K.

APPLICANT: Ahrens, Jeffrey E.

APPLICANT: Ball, James A.

APPLICANT: Bann, G.

APPLICANT: Bell, Erin

APPLICANT: Boddupalli, Raghava

APPLICANT: Deikman, Jill

APPLICANT: Deng, Molian

APPLICANT: Dong, Jinzhao

APPLICANT: Duff, Stephen M.

APPLICANT: Galligan, Meghan M.

APPLICANT: Hinchey, Brenda S.

APPLICANT: Huang, Shihshieh

APPLICANT: Johnson, G. Richard

APPLICANT: Jung, Vincent

APPLICANT: Kretzmer, Keith A

APPLICANT: Laccetti, Lucille B.

APPLICANT: Lai, Chao-Qiang

APPLICANT: Lee, Gary

APPLICANT: Lin, Jie-Yi

APPLICANT: Liu, Jingdong

APPLICANT: Lu, Bin

APPLICANT: Luethy, Michael M.

APPLICANT: Lund, Adrian

APPLICANT: Madson, Linda L.

APPLICANT: Malloy, Kathleen A.

APPLICANT: McKiel, Christine L.

APPLICANT: Miller, Philip W.

APPLICANT: Padmavathi, Manchikanti

APPLICANT: Parnell, Laurence D.
 APPLICANT: Start, William G.
 APPLICANT: Tennessee, Dan

APPLICANT: Vidya, K.R.
 APPLICANT: Wang, Haiyun
 APPLICANT: Xin, Zhanguo

APPLICANT: Xu, Nanfei
 APPLICANT: Yang, Chunzhi
 APPLICANT: Zeng, Xiaoping

APPLICANT: Zhang, Qiang
 APPLICANT: Zhao, Yajuan
 APPLICANT: Zhou, Li

TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants

FILE REFERENCE: 38-15(52796)B

CURRENT APPLICATION NUMBER: US/10/310,154

CURRENT FILING DATE: 2002-12-04

PRIOR APPLICATION NUMBER: 60/337,358

PRIOR FILING DATE: 2001-12-04

NUMBER OF SEQ ID NOS: 736

SEQ ID NO 481

LENGTH: 912

TYPE: PRT

ORGANISM: Zea mays

10-310-154-481

Query Match 85.4%; Score 3883.5; DB 12; Length 912;

Best Local Similarity 84.6%; Pred. No. 2.7e-268;

Matches 772; Conservative 76; Mismatches 58; Indels 7; Gaps 6;

1 MNPEKTHKNTTATAHELAVNAGHAQFTPLHLAGALISDPTGIFFOAISSA-GGENAA 59

1 MNPDNFTHKNEAIVGAHEIAVEAGHAQFTPLHAAVLAADKGGILRAITGASGGDGA 60

60 -QSERVINOALKKLPSPPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQIMGLL 118

61 GDSFVUNSLKLLPSPPDVPASTALIKVIRRAQAQKSGDTHLAVDQIMGLL 120

119 EDSQIRDLNNEVGATARKVEKLEKGEKGVESASGDTNFQALTYGRDLVEQAGKL 178

121 EDSQISDLCKEAGVSAARVRALEKLRGEGRRVESASGDTNFQALTYGRDLVEQAGKL 180

179 DPTVGRDEIRRVIRLRRRTKNNPVLIGEGVGKTAVVEGLAQRIKGVDPNSLTVRL 238

181 DPTVGRDEIRRVIRLRRRTKNNPVLIGEGVGKTAVVEGLAQRIKGVDPNSLTVRL 240

239 ISLDMGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDEIHLVLAGKTEGSDAAN 298

241 IALDMGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDEIHLVLAGKTEGSDAAN 300

239 LFKPMLARGQLRCIGATTLEEYRKVEKDAAFERRFQOVVVAEPSVPTTISILRGLKEKY 358

301 LFKPMLARGQLRCIGATTLEEYRKVEKDAAFERRFQOVVVAEPSVPTTISILRGLKEKY 360

359 EGHGVRVQDRALINAAQISARYITGRHLPKADTLVDDEACANVRVQLDSQPEIDNLER 418

361 EGHGVRVQDRALINAAQISARYITGRHLPKADTLVDDEACANVRVQLDSQPEIDNLER 420

419 KRMQLEIELHALEREKDKASKARLIEVRKEKLDLRLDKLQPLTMKYRKEKERIDIRLQK 478

421 KRIQLELVELHALEREKDKASKARLIEVRKEKLDLRLDKLQPLTMKYRKEKERIDIRLQK 480

479 KREELMFSIQEAREYDILARAADLYGAIQVESIAIQLETSSEENVMLTENVPPIHIA 538

481 RREELQFTIQEAREYDILARAADLYGAIQVESIAIQLETSSEENVMLTENVPPIHIA 539

539 EYVSRWTGTPVTRLQONKEKERLIGLADRLHKKRVVQONAVNAVSEAILRSRAGLGRQOP 598

540 EYVSRWTGTPVTRLQONKEKERLIGLADRLHKKRVVQONAVNAVSEAILRSRAGLGRQOP 599

599 TGSFLLFGPTGVGKTELAKALAEQLFDDENLLVRIDMSEYMEQHSVRLIGAPPYVGHE 658

600 TGSFLLFGPTGVGKTELAKALAEQLFDDENLLVRIDMSEYMEQHSVRLIGAPPYVGHE 659

659 EGGQLTEAVRRRPPCYVILFDEVEKARVAVFNTLLQVLDGRLTDGQRTVDFRNSVILMT 718
 660 EGGQLTEAVRRRPPCYVILFDEVEKARVAVFNTLLQVLDGRLTDGQRTVDFRNSVILMT 719
 719 SNLGAHLLAGLTGKVTMEVARDCVNREVRKHFRPELLNRLDEIVFDPPLSHDQLRKVAR 778
 720 SNLGAHLLAGLTGKVTMEVARDCVNREVRKHFRPELLNRLDEIVFDPPLSHDQLRKVAR 779
 779 LQMKOVAVLAERGVVALAVTDAALDYILAESVDPVYGARPIRRWMEKKVVTLSKMWVRE 838
 780 LQMKOVAVLAERGVVALAVTDAALDYILAESVDPVYGARPIRRWMEKKVVTLSKMWVRE 839
 839 EIDENSTVYIDAGAG--DIVYRVE--SGGLVDASTGKSDVLIHANGPKRSDAAQAVKQM 895
 840 EIDENCTVYIDAGAG--DIVYRVE--SGGLVDASTGKSDVLIHANGPKRSDAAQAVKQM 899
 896 RI-EEEDDDNEE 907
 900 RIMEDEDEGMDEE 912

RESULT 2

US-10-369-493-20047

; Sequence 20047, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 20047

; LENGTH: 881

; TYPE: PRT

; ORGANISM: No. US20030233675A1loc punctiforme

US-10-369-493-20047

Query Match 50.9%; Score 2315; DB 12; Length 881;
 Best Local Similarity 53.9%; Pred. No. 2.1e-156;
 Matches 467; Conservative 172; Mismatches 199; Indels 28; Gaps 11;

2 NPEKTHKNTTATAHELAVNAGHAQFTPLHLAGALISDPTGIFFOAISSAGGENAAQS 61

12 NPEQTEKAWALVTPETAKQFQHQIESEHMLALL-EQEGE-ASSIFNKAGVNVQKL 69

62 AERVINOALKLPSQSPDDIPASSSLIKVIRRAQAQKSGDTHLAVDQIMGLLSDS 121

70 HERTIDFIRNPQKVGSGSSGVIIGHSLERLDRAEQYKKEFGDEYIISIEHLILAFADK 129

122 QI-RDLLNVEVGATARKVEKLEKGEKGVESASGDTNFQALTYGRDLVEQA--GKL 178

130 RFGKGLFQEGDLEKRLNIQQIRGQ--KVTQNPVKVEALEKYGRLTOLAHEGIL 187

179 DPTVGRDEIRRVIRLRRRTKNNPVLIGEGVGKTAVVEGLAQRIKGVDPNSL----- 233

188 DPTVGRDEIRRVIRLRRRTKNNPVLIGEGVGKTAVVEGLAQRIKGVDPNSL----- 247

234 --TDVRLISLDMGALVAGAKYGEFEERLKSVLKEVEDAEKGVILFIDEIHLVLAGKTE 291

248 GRHRKLLIADMALTAGAKYGEFEERLKSVLKEVEDAEKGVILFIDEIHLVLAGKTE 307

292 GSDMAANLFPKMLARGQLRCIGATTLEEYRKVEKDAAFERRFQOVVVAEPSVPTTISIL 351

308 GSDMAANLFPKMLARGQLRCIGATTLEEYRKVEKDAAFERRFQOVVVAEPSVPTTISIL 367

	Query Match	50.3%	Score 2288;	DB 12;	Length 872;
	Best Local Similarity	52.6%	Pred. No. 1.7e-154		
	Matches 460;	Conservative 167;	Mismatches 259;	Indels 18;	Gaps 7
QY	2	NPEKTHKVNETATAHELAVNAGHQAFTPLHAGALISDPGTGIPQAAISSAGGENNAQS	61		
				
Db	5	DNKFTKEAWEAIAKTPRIAKHQHQOIEHLHSALL-EONGIATSIENKAGA----	58		

597 LSDPNRPTASFIPLGPTGVGKTELAKALAKNLDTEALVRLDMSEYMEKHAVALRMGAP 656
 652 PGYVHEEGQOLTEAVRRPYPYCVILFDEVEKAHVAVENTLLQVLDGRLTDGQGRVDFR 711
 657 PGYVHEEGQOLTEAVRRPYPYCVILFDEVEKAHVAVENTLLQVLDGRLTDGQGRVDFR 716
 712 NSVIIMTSLGAEHLIAGLTGKVTMEVARDCVNRVRRKHFPELNLRLDEIVVDFPLSHD 771
 717 NTIIIMTSLGAEHLIAGLTGKVTMEVARDCVNRVRRKHFPELNLRLDEIVVDFPLSHD 776
 772 QLRKVARLOMADVAVRLAERGVAVLTDAAALDYILAESYDVPVYGPPIRRMWEKKVVTTEL 831
 777 ELRSIVQIQSLATLEBQKLTLDKALDFLAAGVDPVYGPPIRRMWEKKVVTTEL 836
 832 SKOVVREIDENSTVYIDAGADLVYRVESGGIV 865
 837 AKGILRGDYKPGETIVVDETDRLSFTSLRGDLV 870

RESULT 4
 ; 10-369-493-2762
 ; Sequence 2762, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 2762
 ; LENGTH: 872
 ; TYPE: PRT
 ; ORGANISM: Synechocystis sp.
 ; 10-369-493-2762

Query Match 50.3%; Score 2288; DB 12; Length 872;
 Best Local Similarity 52.6%; Pred. No. 1.7e-154;
 Matches 460; Conservative 167; Mismatches 229; Indels 18; Gaps 7;
 2 NPEKFTKNTETIATAHVAHAGHAQTPPLHLAGALISDPTGIFPQAISAGGENAQAQS 61
 5 DPNKTEKAWAIAKTPEIAKHRRQQOQIETHEHLLSALL-EQNGLATSIFFKAGA-----S 58
 62 AERV---INQALKKLPSQSPDDIPASSSLIKVIRRAQAQKSGRGTDLAVDQIMGLL 118
 59 IPRVNDQVNSFTAQPKLSNPSESYLGRSLDKLDAEIAKSKYGDYISIEHLMAYG 118
 119 EDSQI-RDLLNEGVATARVSEVKLRGKGGKVESAGDTNFQALTYGRDLVQQA-- 175
 119 QDRGLKGLNYREIGLTENKLAIEIKQIRGTQ--KVTDQNPGEKYESLEKYGRDLTLEARE 176
 176 GKLDVIGRDEIRRVRIILSRRTKNNPVLIGEPGVGTATVEGLAQRIYKGDYVNSLTD 235
 177 GKLDVIGRDEIRVTIQILSRRTKNNPVLIGEPGVGTATVEGLAQRIYKGDYVNSLTD 236
 236 VRLISLDGVALVAGYRGEFEERLKSVLKEVEDAEGKVLIFIDEIHLVLAGKTEGSM 295
 237 RKLISLDGVALVAGYRGEFEERLKSVLKEVEDAEGKVLIFIDEIHLVLAGKTEGSM 296
 296 AANLKPMALRGALRCIGATILEYRKVKVEKDAAFERRPQVYVAPSPVPTISILGLK 355
 297 AGNLLKPMALRGALRCIGATILEYRKVKVEKDAAFERRPQVYVAPSPVPTISILGLK 356
 356 EYVEGHGVTRQDALINAAQLSARYINGRHLPKADLDVDEACANVRVQLDSQPEIDN 415

Db QY 357 ERYVEHGVKIADSAIAAAMLSNRYISDRFLPKAIDLVDAAAKLMEITSKPELDE 416
 416 LERKBMQLEIELHALEREKDKASKARLIEVRKELDDLRDLKQLPTMKYRKEKERIDEIRR 475
 417 VDRKIQLEMERISLQRENDASAKERLEKLEKLAADFKEQSKINGOWQSEKTVIQIRT 476
 476 LKQKRELFMSLOAERYDLARAADLRVGAIOEVRSATIAQLRGTSSEENV---MLTEN 531
 477 VKETIDQVNLIEIQQAQEDVDYNAAELOYGKLTDLQROVEALQTLAQOOTSXSLREE 536
 532 VGPEHIAVSVRGTGIPVTRLGQNEKERLIGLADRLHKKRVVQGNQAVNAVSEAILRSRAG 591
 537 VLSDIAEIIISKVTGIPISKLVSEKEKLLHLEDELHRSRVIGQDEAVTAVAEAIQSRAG 596
 592 LGRQAQPTSGFLPGTGVGKTELAKALAEOLPDDENLLVRLDMSEYMEKHAVALRMGAP 651
 597 LSDPNRPTASFIPLGPTGVGKTELAKALAKNLDTEALVRLDMSEYMEKHAVALRMGAP 656
 652 PGYVHEEGQOLTEAVRRPYPYCVILFDEVEKAHVAVENTLLQVLDGRLTDGQGRVDFR 711
 657 PGYVHEEGQOLTEAVRRPYPYCVILFDEVEKAHVAVENTLLQVLDGRLTDGQGRVDFR 716
 712 NSVIIMTSLGAEHLIAGLTGKVTMEVARDCVNRVRRKHFPELNLRLDEIVVDFPLSHD 771
 717 NTIIIMTSLGAEHLIAGLTGKVTMEVARDCVNRVRRKHFPELNLRLDEIVVDFPLSHD 776
 772 QLRKVARLOMADVAVRLAERGVAVLTDAAALDYILAESYDVPVYGPPIRRMWEKKVVTTEL 831
 777 ELRSIVQIQSLATLEBQKLTLDKALDFLAAGVDPVYGPPIRRMWEKKVVTTEL 836
 832 SKOVVREIDENSTVYIDAGADLVYRVESGGIV 865
 837 AKGILRGDYKPGETIVVDETDRLSFTSLRGDLV 870

RESULT 5
 ; 10-047-260-38
 ; Sequence 38, Application US/10047260
 ; Publication No. US20020164706A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Huang, Lisa
 ; APPLICANT: McClusky, Michael
 ; APPLICANT: Larossa, Robert
 ; TITLE OF INVENTION: HIGH LEVEL PROMOTERS FROM CYANOBACTERIA
 ; FILE REFERENCE: CL1715 US NA
 ; CURRENT APPLICATION NUMBER: US/10/047,260
 ; CURRENT FILING DATE: 2002-01-15
 ; PRIOR APPLICATION NUMBER: 60/264,925
 ; PRIOR FILING DATE: 2001-01-30
 ; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 38
 ; LENGTH: 872
 ; TYPE: PRT
 ; ORGANISM: Synechocystis sp. strain PCC6803
 ; 10-047-260-38

Query Match 50.3%; Score 2288; DB 14; Length 872;
 Best Local Similarity 52.6%; Pred. No. 1.7e-154;
 Matches 460; Conservative 167; Mismatches 229; Indels 18; Gaps 7;
 2 NPEKFTKNTETIATAHVAHAGHAQTPPLHLAGALISDPTGIFPQAISAGGENAQAQS 61
 5 DPNKTEKAWAIAKTPEIAKHRRQQOQIETHEHLLSALL-EQNGLATSIFFKAGA-----S 58
 62 AERV---INQALKKLPSQSPDDIPASSSLIKVIRRAQAQKSGRGTDLAVDQIMGLL 118
 59 IPRVNDQVNSFTAQPKLSNPSESYLGRSLDKLDAEIAKSKYGDYISIEHLMAYG 118
 119 EDSQI-RDLLNEGVATARVSEVKLRGKGGKVESAGDTNFQALTYGRDLVQQA-- 175
 119 QDRGLKGLNYREIGLTENKLAIEIKQIRGTQ--KVTDQNPGEKYESLEKYGRDLTLEARE 176

JLT 6
 00-10-156-761-14770
 Sequence 14770, Application US/10156761
 Publication No. US20030119018A1
 GENERAL INFORMATION:
 APPLICANT: OMURA, SATOSHI
 APPLICANT: IKEDA, HARUO
 APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHIYUKI
 APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCES: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 CURRENT FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109
 SEQ ID NO 14770
 LENGTH: 879
 TYPE: PRT

APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 20488
 LENGTH: 857
 TYPE: PRT
 ORGANISM: Rhodospirillum rubrum
 10-369-493-20488

Query Match 49.8%; Score 2264.5; DB 12; Length 857;
 Best Local Similarity 53.1%; Pred. No. 8e-153;
 Matches 458; Conservative 159; Mismatches 234; Indels 11; Gaps 7;

1 MNPEKFTHTNETIATATHELAVNAGHAQFTPLHLAGALISDPTGIFPQAISAGGENAAQ 60
 1 MNVEKTERVGFQSAQSLARWEGHQFSPHLHKLVDSDSEGLAGLIDRAGNSRA- 59
 61 SAERVINQALKPLSQS--PPDDIPASSSLIKVIRRAQAOKSRGDTHLAVDQIMGLL- 118
 60 -ILKATEALGKMPKVSQSGAGQVYLPATAPDAEKAEEKAGDSFVTVRLLALLSL 118
 119 -EDSQIRDLNVEGVATARKVSEKVLKRGKGVKVESAGDTNFQALKTGYGRDILVEQA-- 175
 119 DKDSAGQLTKGGVTPQNLNAINALR--KGRADSATAENAYDALKKYARDLTQARD 176
 176 GKLDPVIGRDEIRRVIRLSRRTKNNPVLIGEPGVGTAVVEGLAQRIKGVDPNSLTD 235
 177 GKLDPVIGRDEIRRVIRLSRRTKNNPVLIGEPGVGTAVVEGLAQRIKGVDPNSLTD 236
 236 VRLISLDMGALVAGAKYRGEPEERLKSVLKEVEDAEAGKVLFIIDEIHLVLAGKTEGSM 295
 237 KKLALDMGALLIAGAKYRGEPEERLKSVLKEVEDAEAGKVLFIIDEIHLVLAGKTEGSM 296
 296 AANLKPMLARGQLRCIGATTLEEVKVKVEKDAAPERRPQQVYVAESPVTITSLRGLK 355
 297 ASNLKPALARGELHCIGATTLEEVKVKVEKDAALARRFPQVFPVSEPTVEDTVSLRGLK 356
 356 EYEGHGVRIQDRALINAAQLSARYITGRHLPKADILVDEACANVRVQLDSOPEEIDN 415
 357 DKYEGHGVRIQDRALINAAQLSARYITGRHLPKADILVDEACANVRVQLDSOPEEIDN 416
 416 LERKMQLEIEIHALEREKDKASKARLIEVRKELDDLRDLPKADILVDEACANVRVQLDSOPEEIDN 475
 417 MDREIVRLKIEQALKKETTDPGSKARLVLEKELADLEKSAALTORWSAEKNKLSAQK 476
 476 LKQREELMFSLOEABERYDIARAADLYGAIQEVESAIAOLEGTSSEENVMLTENVGPE 535
 477 LKSELDALRIELANAGRGYQAGELAYGRIPLEKXIAIE-ANENSGAMVEAVTAN 535
 536 HIAEVSWTGTIPVTRIGQNKERLIGIADRLHKKRVGQONAVNAVEATILRSAGLAGRA 595
 536 HIAQVVSRTGVPVDMKLEGEKELRMBEQGQVVGQFEAVHASTAVARRAGLQDP 595
 596 QOFTGSFLFGPTGVKTELAKALAEQFDDENLLVRIDMSYEMQHSVSLRIGAPFGYV 655
 596 NRPMGSEFNLGPTGVKTELAKALAEYFDDETAMVRIDMSEFMEKHSVSLRIGAPFGYV 655
 656 GHEEGQLTEAVRRPYCVILFDEVEKAHVAFTNLLQVLDGRLTDGQRTVDFRNSVI 715
 656 GYDEGGVTEAVRRPYCVILFDEVEKAHVAFTNLLQVLDGRLTDGQRTVDFRNTLI 715
 716 IMTSLNLAGHLLAGLTGKVTMEVARDCVMEVRKHFPELLANRLDEIVDPDLSHDLRK 775

Db 716 VMTSLNGSELYNQPEGEDTGAV-REQVMGVRAHPRPBEFLNRVDEIILFHLRLOKSEMR 774
 QY 776 VARLOMKQVAVRLAERGVAVLAVTDAALDVLIAESYDPVYGARPIRRWMEKKVVTLSKMW 835
 Db 775 IVDIQLARLTKLEDRKIVLDDAARDWIAEKGMDDPAYGARPLKRVORSVDPLAEMI 834
 QY 836 VREEDENSTVYIDAGAGDLVY 857
 Db 835 LEGSVKGDGHVAISAEQGVLT 856

RESULT 8
 US-10-369-493-19762
 ; Sequence 19762, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 19762
 ; LENGTH: 862
 ; TYPE: PRT
 ; ORGANISM: No. US20030233675A1altoc punctiforme
 US-10-369-493-19762

Query Match 49.4%; Score 2247; DB 12; Length 862;
 Best Local Similarity 52.0%; Pred. No. 1.4e-151;
 Matches 446; Conservative 174; Mismatches 220; Indels 18; Gaps 7;

QY 2 NPEKFTHTNETIATATHELAVNAGHAQFTPLHLAGALISD---PTGIFPQAISAGGENA 58
 Db 5 NNQFTEKAWAIAHTPDIVKQYQQQITSEHLMKALLLEQDGLATGILTKA-----GVNL 59
 QY 59 AQAERVINQALKPLSQSPDDIPASSSLIKVIRRAQAOKSRGDTHLAVDQIMGLL 118
 Db 60 QKLRDT-EQFFQRPQKVSSTSSVYLGRLDRLDRAVYKRFQDEYISTEHLLEYA 118
 QY 119 EDSQI-RDLNVEGVATARKVSEKVLKRGKGVKVESAGDTNFQALKTGYGRDILVEQA-- 175
 Db 119 KDRFGKALFQFGLDEGKLDIILKVRGSO--KYTDQNPGEKYLEKYGKRLTEAARK 176
 QY 176 GKLDPVIGRDEIRRVIRLSRRTKNNPVLIGEPGVGTAVVEGLAQRIKGVDPNSLTD 235
 Db 177 GQLDPVIGRDEIRRVIRLSRRTKNNPVLIGEPGVGTAVVEGLAQRIKGVDPNSLTD 236
 QY 236 VRLISLDMGALVAGAKYRGEPEERLKSVLKEVEDAEAGKVLFIIDEIHLVLAGKTEGSM 295
 Db 237 KKLALDMGALLIAGAKYRGEPEERLKSVLKEVEDAEAGKVLFIIDEIHLVLAGKTEGSM 296
 QY 296 AANLKPMLARGQLRCIGATTLEEVKVKVEKDAAPERRPQQVYVAESPVTITSLRGLK 355
 Db 297 AGNLLKPMARGELHCIGATTLEEVKVKVEKDAALARRFPQVYVQPSVDSISILRGLR 356
 QY 356 EYEGHGVRIQDRALINAAQLSARYITGRHLPKADILVDEACANVRVQLDSOPEEIDN 415
 Db 357 ERYENHGVKISDSALVAAVLSRYISDRFLPKADILVDEACANVRVQLDSOPEEIDN 416
 QY 416 LERKMQLEIEIHALEREKDKASKARLIEVRKELDDLRDLPKADILVDEACANVRVQLDSOPEEIDN 475
 Db 417 IDRKILQLEMEKLSLQKESDAASRERLERLEKEIADLKEEQRTLTNTQWSEKDIIDKIQS 476
 QY 476 LKQREELMFSLOEABERYDIARAADLYGAI---QEVESAIAOLEGTSSEENVMLTEN 531

QY 6 FTHKTNETIATAHELAVNAGHAQFTPLHLGALISDP-----TGIFPQAISSAGGENAAQS 61

6 FTDKAAKTVLSAYGIAQSGHSLQTPHIAAALLSDSDSNGTTLRLTRIVDKAGD--GQK 63
 62 AERVINOALKKLPQSPPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLMGLLEDS 121
 64 PERSVTSLSVLPAQDPPPEQVTLSPESAKLLRNAHELOKTKQDSYIAQDHFIAVFTKDD 123
 122 QIRDLLENGVATARVSEVEKLKRGKGVESAGDTNFQALTKYGRDLVEQA--CKLD 179
 124 TLKSLAAGVTPKAFEFANVRG--NKRIDSXNAEEGFDALNKFTVDTLTELARNQOLD 181
 180 PVIGREIRVRVRLSRTKNNPVLIGEPGVGKTAVVGLAQRIVKGDVPSNLTVDRLI 239
 182 PVIGREIRVRVRLSRTKNNPVLIGEPGVGKTSTAEGLARLIIDDDVPANLSNCKLL 241
 240 SLDMGALVAGAKYGEFEERLKSVLKEVEDAEGKVILFIDIEHIVLGAQK--TEGSDMAAN 298
 242 SLDVGSVLAVGSKFRGEPEERIKSVLKEVESETPILFVDEMHLLMGAGSGEGGMDAAN 301
 299 LFKPMARGQLRCIGATTLBEYRKVVEKDAAFERRFOQVYVAEPSVDTTISILGLKEKY 358
 302 LKPMARGQLKHCIGATTLAARYKYKIEKDAAFERRFOQVYVAEPSVDTTISILGLKEKY 361
 359 EGHGVRIQDRALINAAQLSARYITGRHLDPKALDLDVDEACANVRVOLDSDPEEIDMLER 418
 362 EVHGVVTLSDRALVTAHLASRYLTSRLPDSALDLDVDEAAAARVTVRESQPEVLDNLR 421
 419 KRMQLEIELHALEREKOKASKARLIEVKELEDDLRDKLOPLTMKYRKEKERIDEIRLKQ 478
 422 KLRQVRVEIRALEREKDEASKERLKAARKEAEQVEEETPRIRREYKELEKSRGSELOAKR 481
 479 KREELMFSLOEAREVYDLARADLRGYGAIQVESAIQOEGTSSEENV----- 526
 482 RLDELKAKAEDARRNDFTLADLKYYGIPDQKRIEYLEQOKKADABAFANAQPSSEP 541
 527 MLTENVGPEHIAEVVSRWTGIPVTRLGONEKERLIGLADRLHKRVVQGNQAVNAVSEAIL 586
 542 LLIDVVGPDQINEIVARWTGIPVTRLTKTEKERLUNMEKVLKQVIGQNEAVTAVNAIR 601
 587 RSRAGLGRAQQTGSLFLGPTGVGKTELAKALABQLFDDENLLVRIDMSYMEQHSVSR 646
 602 LSRAGLSPNQPIASFPCGSPSGTKTLTKALASFDFDENAMIRIDMSYNEKHSVSR 661
 647 LIGAPGVGHEEGQLTAEVRRPYCVILFDEVEKAVHVFNTLLVLDGRLTDGGR 706
 662 LIGAPGVGHEAGQLTQLRRRPYVILFDEIEKAAPEVTLVLLVLDGRLTSGGQ 721
 707 TVDERSVITMTSNLGAELLA--GLTKVTVMEVARDVMEVRKHPPELLNRLDEIV 763
 722 VVDKNAVILMTSNLGAELVITDNESSDDGKID-STTREMVMNSIRGFFRPFELNRSIV 780
 764 VPDPLSHDLQRLKVARLQMDVAVRLA--ERGVALAVTDAALDYILAEYDVPYGNARPIR 821
 781 IENRRLVRDIRNIVENRILEVQKRLQSNHRSIKIEVSDEAKDLGSGAGSYPAYGARPLNR 840
 822 WMEKQVTELSKVMREEIDENSTVYIDAGADLVVRVESGGLVDASTGCKSDVLIHLIAN 881
 841 VTQNVLPNMAVLIINGQLRDKETAHVVVQNGKI FVKN-----HEAN 883
 882 GPKRSDAAQAVKVMRIEIEDDDNEEMIE 910
 884 ANGSAD-----IDMDGIDDDVNDDELE 905

SULT 11

10-369-493-11982
 Sequence 11982, Application US/10369493
 Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 11982
 LENGTH: 868
 TYPE: PRT
 ORGANISM: Mesorhizobium loti
 US-10-369-493-11982

Query Match 48.9%; Score 2225; DB 12; Length 868;
 Best Local Similarity 52.0%; Pred. No. 5.3e-150;
 Matches 448; Conservative 159; Mismatches 245; Indels 10; Gaps 6;

QY 1 MNPEKTHKTNETIATAHELAVNAGHAQFTPLHLAGALISDPTGIFPOAISAGGENAAQ 60
 DB 1 MULEKISERVGFQISQAQTMALSRNHQFTPEHILKVLVDDDEGLAASLIERAGN--VR 58
 QY 61 SAERVINOALKKLPQSPPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLMGLL-- 118
 DB 59 DVKLGVTETALEAMPKVEGNGQQLYLAQPLAKVFSTAEELAKKAGDSFVTVERLLQALTWE 118
 QY 119 EDSQIRDLLENGVATARVSEVEKLKRGKGVESAGDTNFQALTKYGRDLV--EOAG 176
 DB 119 KSAKTADIIAKAGVTAQALNQVINDVR--KGTADASABQNTDALKKYARDLTADARAG 176
 QY 177 KLDPVIGRDEEIRRVVRLSRTKNNPVLIGEPGVGKTAVVGLAQRIVKGDVPSNLTVD 236
 DB 177 KLDPVIGRDEEIRRTIQLVLSRRTKNNPVLIGEPGVGKTALVAGLALRVNGDVPSLKDOK 236
 QY 237 RLISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEGKVILFIDEHLVLGAGKTEGSDMA 296
 DB 237 QLMALDMGALLIAGAKYRGEFEERLKAVALSEVTSANGNIILFIDEMHTLVGAKGADGAMDA 296
 QY 297 ANLKPMLARGQLRCIGATTLBEYRKVVEKDAAFERRFOQVYVAEPSVDTTISILRGLKE 356
 DB 297 SNLLKPKALARGELHCVGATTLDEYRKHVEKDPALARKFQVFDVETVTSILRGLKE 356
 QY 357 KYEGHGVRIQDRALINAAQLSARYITGRHLDPKALDLDVDEACANVRVOLDSDPEEIDML 416
 DB 357 KYEQHKKVRLSDSALVAATLSNRYIADRFDPKALDLDVDEAASRLRMQVDSKPEALDEI 416
 QY 417 ERKMQLEIELHALEREKOKASKARLIEVKELEDDLRDKLOPLTMKYRKEKERIDEIRL 476
 DB 417 DRRIMQKIEREALKVETDDASKORLVRLEKELVGLSEESTETAKWQAEKOKLGLAADL 476
 QY 477 KOKREELMFSLOEAREVYDLARADLRGYGAIQVESAIQOEGTSSEENVMLTENVGPEH 536
 DB 477 KQOLDEARNELATAQRKGFQORAGELAYGKIPELEKKLXEAQADGKAG--WVEVVTDPH 535
 QY 537 IAEVSRWTGIPVTRLGONEKERLIGLADRLHKRVVQGNQAVNAVSEAILRSRAGLGRAQ 596
 DB 536 VAHIVSRWTGIPVDMKQLQERDKLRLMEDEIGKRVVVGQGEAVQAVSKAVRRARAGLDQEN 595
 QY 597 OPTGSLFLGPTGVGKTELAKALABQLFDDENLLVRIDMSYMEQHSVSRLLIGAPGVYG 656
 DB 596 RPTGSLFLGPTGVGKTELAKALASFLFDDDSANVRIDMSYMEQHSVSRLLIGAPGVYG 655
 QY 657 HEEGGQLTAEVRRRPYCVILFDEVEKAVHVFNTLLVLDGRLTDGGRITVDYFRNSVII 716
 DB 656 YEEGALTEAVRRRPYQVVLFDIEIKAPDVFVNVLVLDGRLTDGGRITVDYFRNTLII 715
 QY 717 MTSNLGAELHLAGLTKVTVMEVARDVMEVRKHPPELLNRLDEIVVFPDPLSHDLQRLKV 776
 DB 716 MTSNLGAELV--NLGEGDQVDVAVRDEVMGVVRASFRPEFLNRVDEVILFHLRERKMDRI 774
 QY 777 ARLOMKDVAVRLAERGVAVTDAALDYILAEYDVPYGNARPIRREMEKQVTELSKVMV 836
 DB 775 VEIQKRLSLLVDRKITLSLDEHEALEWLAAGKIDPAYGARPLKRVNQKELQDPLAEKIL 834

837 REBIDENSTVYIDAGAGDLYV 858
 835 LGEILDGSIKVTSGSDRLNFR 856

BLT 12
 0-369-493-16834
 quence 16834, Application US/10369493
 blication No. US20030233675A1

GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 PRIOR FILING DATE: 2003-02-28
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 16834
 LENGTH: 859
 TYPE: PRT
 ORGANISM: Caulobacter crescentus
 0-369-493-16834

Query Match 48.9%; Score 2224; DB 12; Length 859;
 Best Local Similarity 51.9%; Pred. No. 6.2e-150;
 Matches 446; Conservative 158; Mismatches 246; Indels 10; Gaps 6;

1 MNPEKTHKNTIATHELAHVAGHAQFTPLHAGNISDPTGIPQALSSAGGNAQ 60
 1 MNIDLYSDRAKQAVQSAQSLAARGHQFAPHEILKVLLEKDGSLALIQSAGGR--PD 58
 61 SAERVINOALKKLPSPQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQIMGLL-E 119
 59 QLDGGVETLAKTPRVGAGGQLYMKPDTARVFAEAKSAKAAGADAPVITTELLIAIKE 118
 120 DSQIRDLLNEVGAVATARKVEKELRGKKEKVESAGDTNFQALKYTGRLVQEA--GK 177
 119 GGEAAKLFKEAGVSAQSLAANAAMR--KGTADSAANAEGYEAALKRYARDLTAARADGK 176
 178 LDPVIGDEIRVRVRLSRRTKNNPVLIGSPGVGTAVVGLAQRIKGVDPNSLTDVR 237
 177 LDPVIGDEIRRTIQVLSRRTKNNPVLIGSPGVGTAVVGLALRVNGDVPESLKDKK 236
 238 LLSIDMGALVAGAKYRGEFEERLKSVLKEVEDAEKGVILFIDEHLVLGAGKTEGMDAA 297
 237 LLSIDMGSLAGAKYRGEFEERLKAVALGEVTAABGSILFIDEMHTLVGAGKGDGMDAS 296
 298 NLFPMLARGQLRCIGATTLEBYKYVEKDAAFERRRQOQYVVAEPSPVDTISILRLKEK 357
 297 NLLKPALARGELHCVGATTLDVEYKHKVDAALARRFPQVFPVSEPTVEDTVSLIRLKEK 356
 358 YEGHGVRIQDRALINAQLSARVITGRHLPDKAIDLVDDEACANVRVOLDSPQEEIDNLE 417
 357 YEVHGVRISSAIVAAATSNRVIADRFDPDKAIDLVDDEASRVMOIDSKPEELDEID 416
 418 RKRQLETELHALEREKOKASKARLIEVRKELDLDRKLOPLTMKYRKEKERIDEIRRLK 477
 417 RRLVQLKTERALSKEITDAASKORLENLEVIDDLQFRSDEMTRAKWEKVKVGAQAR 476
 478 QKRLEMFSLQEAERRYDLAARDLRYGAIQEVESATIAQLGTSSENVMJTNVNGPEHI 537
 477 EALDRRLADLANAQRAGDFARAGQIQYGEIPALERRLAEE--AGDTQALTPEVVDQEI 534
 538 AEVSVRWGTIPVTRLGQNEKERLLGLADRLHVRVQGVQNAVNSEALLSRAGLGRAQQ 597
 535 AAVSVRWGTGVPEVKLEGEREKLKMEDELRGRVVGQDEALEAVSDAVRRARAGLQDPK 594

QY 598 PTGSFLFGPTGVGKTELAKALABQLFDENLLVRIDMSYMEQHSVSRLLGAPPGVGH 657
 DB 595 PIGSFLFGPTGVGKTELAKALABQLFDENLLVRIDMSYMEQHSVSRLLGAPPGVGH 654
 QY 658 EGGQQLTEAVRRPVCVILFDEVEKAHVAVNTLLOVLDGRLTDGQRTVDFNSVLIM 717
 DB 655 DEGGALTEAIRRPVQVVLFEIEKAHPDVFENVLLQVLDGRLTDGQRTVDFNTLLIM 714
 QY 718 TSNLGAELHLAGLTKGVTMEVARDCWREVKHFRPELLNRLDEIVVDFPLSHDQLARKVA 777
 DB 715 TSNLGAELYASQEDGE--DVEAVRPMWMTVGHFRPEFLNRLDEILILFKLSRHNMGDIV 773
 QY 778 RLQMKDVAVRLAERGVAVLAVTDAALDYLAESYDPVYGARPIRRMMKVKVVTLSKMMVR 837
 DB 774 RIQLQVREKLLADRRMALDAEALNLWADKGYDPVYGARPLKRVIQKELVDPIAKKLLA 833
 QY 838 REIDENSTVYIDAGAGDLYV 857
 DB 834 GEIEDGGVIAVGVTGQQLSF 853

RESULT 13

US-10-369-493-3990
 ; Sequence 3990, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 3990
 LENGTH: 907
 TYPE: PRT
 ORGANISM: Neurospora crassa
 US-10-369-493-3990

Query Match 48.7%; Score 2214.5; DB 12; Length 907;
 Best Local Similarity 49.6%; Pred. No. 3.2e-149;
 Matches 468; Conservative 156; Mismatches 240; Indels 79; Gaps 16;

QY 6 FTHKNTIATHELAHVAGHAQFTPLHAGNISDPTGIPQAL 50
 DB 1 FTDRAKALEDAALAEQVAHSQLLPVHLAVALLDPLPDPKDDQONAPAGATSSLFQVI 60
 QY 51 SSAGGENAQAASERVINOALKKLPSPQPPDDIPASSSLIKVIRRAQAQKSRGDTHLAV 110
 DB 61 ERAHGD--PQFDRAKALVRLPSQDPPDHVSNAFHTVTLKAKNELQTKQDTYIAV 118
 QY 111 DOLIMGLLDSQIRDLLNEVGAVATAR--VKSEVEKELRGKKEKVESASGDT--NFOALKTY 167
 DB 119 DDLITLALAEPSIMMALKEANIIPKLVITDAIQARGT--KXVDSRNADTEEHENLAKF 176
 QY 168 GRDLVQEA--GKLDPVIGDEIRVRVRLSRRTKNNPVLIGSPGVGTAVVGLAQRIV 225
 DB 177 TIDMTAAREGKIDPVGREEEIRRVIRLSRRTKNNPVLIGSPGVGTAVVGLAQRIV 236
 QY 226 KGVDPNSLTDVRLISIDMGALVAGAKYRGEFEERLKSVLKEVEDAEKGVILFIDEHLVL 285
 DB 237 NADVPDNLANCKLLSLDVGALVAGKYRGEFEERLKSVLKEVEDAEKGVILFIDEHLVL 296
 QY 286 GAGKT--EGSMDAANLKPMLARGQLRCIGATTLEBYKYVEKDAAFERRRQOQYVVAEPSV 344
 DB 297 GAGAGGEGMDAANLKPMLARGQLRCIGATTLEBYKYVEKDAAFERRRQOQYVVAEPSV 356

```
345 PDIISILRLGKEVGHGVRIOADRALINAAQSARYITGRHLPDKAIDLVDACANVRV 404
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 SETISILRLGKEVGHGVTISDAIVAANLAARYLTSRUPDSALIDLIDEAAAVRV 416
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405 QLDSPQIDNLERKRWQLETELHALEREKDKASKARLIEVRKELDLRLKLOPLTKYR 464
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
417 ARESQPEIISLERKRLQKLEIHALSREKDEASKARLEQAKKDAENVEELRLEKYE 476
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
465 KEKERIDEIRLKKRELMESLOEARRRYDLARAADLRVGAIOEVESALIAOLE----- 518
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
477 QEKORAKALQARWKLESRLQKAEASRMGDHSRAADLOYYAIPQEAIVKRIEKEKAAA 536
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
519 -----GTSSEENVMLTENVGPEPHIAEVVSWTGPVTRLGONKEREKRLIGLADRLHKRVV 572
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
537 DAALNAAAETGGAMITDVVGPQINEIVASWTGPVTRLKTSEKEKLLHMEKLSKIIV 596
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
573 GONQAVNAVSEALRSRAGLQRAOQPGSFLFGPTGVGKTELAKALAEOLFDDENLLVR 632
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
597 GQKEAVQSVNAIRLQRSGLSNPNQPP-SFLFCGSPSGTGKTLTKALAEFLFDDPKAMIR 655
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
633 IDMSYMECHSVSRLLIGAPPGVVGHEGGQLEAVRRPYCVILFDEVEKAHVAVFNTLL 692
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
656 FMSYQERHLSRSMIGAPPGVVGHDGSGQLEALRRKPFSSILLFDEVEKAKEVILVLL 715
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
693 QVLDDGRLTDGQRTVDFRNSVIMTSNLGAHLLA--GLTGKVTMEVARDCVNREVRKH 750
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
716 QLMDDGRLTDGQRTVDFRNSVIMTSNLGAHLLA--GLTGKVTMEVARDCVNREVRKH 774
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
751 FPELNLRLDEIVDPDLSHDLKRVARLQKDVAVRLA--ERGVAVAVTDAALDYILAE 808
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
775 FLPEFLNRISIVIFNRITRREIRKIVELRATQRLQDNRNVKIEVSEAKDKLGA 834
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
809 SYDPVYGARPIRNMKEKVVWVLSKMWVREEDENSTVVIDAGDLVYRVESGGLVDAS 868
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
835 GUSPAYGARPLORVLEKEVNLAVLIIRGSRD-----GEVARVVQDGKITV- 883
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
869 TGKSDVLIHANGPKRSDAAQAVKMKRIBETDDDNEMIED 911
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
884 -----LNPHP-----EVNDEDDENMLDE 901
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

MULT 14
09-815-242-14088
Sequence 14088, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Cart, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308

PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 14088
LENGTH: 861
TYPE: PRT
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(861)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-14088

Query Match 48.4%; Score 2204; DB 9; Length 861;
Best Local Similarity 52.1%; Pred. No. 1.7e-148;
Matches 449; Conservative 155; Mismatches 244; Indels 14; Gaps 9;

Qy 1 MNPEKTHKTNETIATAHELAVNAGHAQPTPLHLAGALISDPTGIFPOAISSAGGENAAQ 60
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5 MRLDRUTNFKQALADADAQSLALGHDNQFTEPLHLSALLNQSGSIRPLLTSA-GINAGQ 63
Qy 61 SAERVINQALKKLPSPDPDDIPASSSLIKVIRRAQAQKSRGDTHTLAVDQIIMGLLED 120
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 -LRTAIDQALSRLPQVGGTGDVQPSSELVRLVNLCDKLAOKRGDNFISSELFVLALES 122
Qy 121 -SQIRDLINVEGVATARVSEVKEKRGKKEKVSASGDTNFOALKTYGRDLVEQA--GK 177
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123 RGTLTLLKSAAGATTANITQAIEMRG--GESVNDQGAEDQOALKKTYVDLTERAQOK 180
Qy 178 LDPVIGRDEIRRVVAILSRRTKNNPVLIGEPGVGKTAIVEGLAQRIYKGVDPNSLTDVR 237
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 LDPVIGRDEIRRTIQVLRRTKNNPVLIGEPGVGKTAIVEGLAQRIYKGVDPNSLTDVR 240
Qy 238 LISLDMGALVAGAKYGEPEERLKSVLKEVEDAEKGVILFIDEIHLVLGAGKTESMDAA 297
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 VLALDMGALVAGAKYGEPEERLKSVLKEVEDAEKGVILFIDEIHLVLGAGKTESMDAA 300
Qy 298 NLPKPMARGOLRCIGATTLEEVKVKVEKDAAPERFQOVVVAEPPVPTTISILRLKKEK 357
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 NMLKPALARGELHCVGATTLDEYRQYIEKDAALERRFKQVFAEFSVEDTIALRLKKEK 360
Qy 358 YEGHGVRIODRALINAAQLSARYITGRHLPDKAIDLVDACANVRVQDLSQPEEIDNLE 417
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 YELHHVQITDPAIVAANTLSHYIADRLQIPDKAIDLIDEAASSIRMQIDSKEPEELRLD 420
Qy 418 RKMQLIEIHLALEREKDKASKARLIEVRKELDLRLKLOPLTKYRKEKERIDEIRLKL 477
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421 RRIITQLKEQALMKESDEASKERLDMLENEELDDKERQYSELEEEKAEKASLSGTQTTK 480
Qy 478 QKBEELMPSLOEARRRYDLARAADLRVGAIOEVESALIAOLE-GTSSEENVN--LTENVGP 534
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 AELEQAKIAEQARRVVDLARMSELOVYKPIPELEK--QLEAATQSEKTMRLLRNKVTD 537
Qy 535 EHIAEVVSRWTGIPVTRLGONKEREKRLIGLADRLHKRVGQONAVNAVSEALRSRAGLGR 594
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
538 AETAELVARTGIPVSRMLSGEREKLLRMQEQLHSRVIGQNEAVNAVSEALRSRAGLSD 597
Qy 595 AQOQTGSLFLGPTGVGKTELAKALAEOLFDDENLLVRIDMSYMECHSVSRLLIGAPPGY 654
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
598 PNRPIGSLFLGPTGVGKTELAKALAEOLFDDENLLVRIDMSYMECHSVSRLLIGAPPGY 657
Qy 655 VGHEGGQLEAVRRPYCVILFDEVEKAHVAVFNTLLQVLDDGRLTDGQRTVDFRNSV 714
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
658 VGEYEGGLTAVRRPYCVILFDEVEKAHVAVFNTLLQVLDDGRLTDGQRTVDFRNSV 717
Qy 715 IIMTSNLGAHLLAGLTGKVTMEVARDCVNREVRKHFRPELNLRLDEIVDPDLSHDLQR 774
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
718 VIMTSNLGSD-LIQERFGLDYGRMKEMVLGVVVSQNFPEFINRIDEVVFVHPLGEQHIA 776
Qy 775 KVARLQKDVAVRLAERGVAVTDAALDYILAEVDPVYGARPIRNMKEKVVVTELKSM 834
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
777 SIAQIQLQRLYKRLREERGVEIHSDEALKLSANGYDVPVYGARPLKRAIQOQIENPLAQO 836
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

835 VVREIDENSTVYIDAGADLV 856
:: : : : :
837 ILSGELVPGKVRLEANDDRIV 858

JUL 15
10-369-493-10608
sequence 10608, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 10608
LENGTH: 858
TYPE: PRT
ORGANISM: Sphingomonas aromaticivorans
10-369-493-10608

Query Match 48.3%; Score 2199.5; DB 12; Length 858;
Best Local Similarity 52.4%; Pred. No. 3.5e-148;
Matches 454; Conservative 160; Mismatches 232; Indels 21; Gaps 9;
1 MNPEKFTHTNETTATAHELAVNAGHAGFTPLHAGALISDPTGIPPOAISSAGGENAAQ 60
1 MNLEKFTDRAGKFPQAAQTVAIRNHHORITPDHILKALLEDESEGMAAGLQACG-NAA- 58
61 SAERVINOALKKLP-----SQSPPPDDIPASSSLIKVIRRAQAQKSRGDTHLAVDQLI 114
59 LAQTEVDKALAKIPAVSGSQAQATPGLD---NDARVLDLSAEQIAAKSNDSFVTVERML 114
115 MG--LLEDSQIRDLLNEVGATARKVEKLRGKQKVESASGDTNFQALKTYGHDLV 172
115 VALTIATTSAGQALKAAANTVAQALEAATQLRG--GRTADSASAEANAYDAMKKYARDLT 172
173 EQA--GKLDPVIGDEERRVRIILSRRTKPNVLI GEPGVGKTA VVEGLAQRIVKGDVP 230
173 EAAREGKLDPVIGDEERRVRIILSRRTKPNVLI GEPGVGKTA VVEGLAQRIVKGDVP 232
231 NSLTDVRLISLDMGALVAGAKYRGFEERLKSVLKEVEDAEGKVLFIIDHILVLGAGKT 290
233 DSLKDRRLMALDMSLITAGAKYRGFEERLKA VLDVKGAEGETILFIDEMHTLIGAGKS 292
291 EGSMDAANLKPMLARGOLRCIGATTLEEYKRYVEKDAAPERFQOYVVAEPSVPTISI 350
293 EGAMDAGNLLKPALARGELHCIGATTLEDEYQYVEKDPALQRRFPQVFGVEPTVEDTISI 352
351 LRGLKEKYEGHGVRIODRALINAQLSARYITGKHL PDKALDLDVDEACANVRVOLDSP 410
353 LRGIKDKVELHHGVRIADNAIVAATLSNRVYISDRFLPDKALDLDVDEAASRIRMEVESKP 412
411 BEIDNLRKRNQLEIELHALREKDKASKARLIEVRKELDDLRKQLQPLTKYRKKEKERI 470
413 EEIEKLDRIIQMKIEENALAKETDQASKDRLATLREELANQEQOQSAELTTRWQNERDKI 472
471 DEIRLQKRELWFSLOEERRVYDARAADLYRGAIOEVESAIAQLEGTSSEENVMLTE 530
473 AAEQKKEALDAARSELEVAQRNGDLKAGAGLAYGRIPELERQLAEAGGV--QNAMLRE 530
531 NVGPEHIAEVVSRWTGIPVTRIGQNEKERLIGLADRLHKRVVQGNQAVNAVSEAILRSRA 590
531 EVTAEDIAAVVSKWTGVPDRMMEGEREKLHMEALGKRVIGQKDAVLAVSKAVRARA 590

QY 591 GLGRAQQPTGSEFLGPTGVGKTELAKALAEQLFDDENLLVRIDMSYMEQHSVSRLLIGA 650
DB 591 GLQDPNRPLGSFLGPTGVGKTELAKALAGFLFDDDNAMVRIDMSYMEQHSVSRLLIGA 650
QY 651 PFGYVCHBEGGQITAVRRRPPYCVILFDEVEKAVAVFNTLLQVLLDDGRLTDGQGRTVDF 710
DB 651 PFGYVGYDEGGVLTETAIRRRPYQVVLFDVEVEKASDVFNVLQVLLDDGRLTDGQGRVVD 710
QY 711 RNSVIIMTSNLGAELHLAGLTGKVTMEVARDCVMEVRKHPPELNLRLDEIVVDFPLSH 770
DB 711 TWTLLILTSNLGSQY-LANLEGGQVQSVPEQVMDVRAHFPPELNLRLDEILFHLGH 769
QY 771 DQLRKVARLQMKDVAVRVAERGVAVLADDAALDYILABSYPVYGARPIRRMMEKVVTE 830
DB 770 EEMAPIVDIQVGRVANLLKDKRIVLDTDAAKRWLGRVGYDPVYGARPLKRAVQRYLQDP 829
QY 831 LSKWVVRBEIDENSTVYIDAGADLVY 857
DB 830 LAEKLLIGGEVDPGTVRIDEGDGLSF 856

Search completed: February 13, 2004, 01:43:03
Job time : 82 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

nucleic - nucleic search, using sw model

on: February 12, 2004, 17:25:21 ; Search time 770 seconds
(without alignments)
10885.396 Million cell updates/sec

le: US-09-812-350-30
fect score: 3105
quence: 1 aaagtatcaattttacac.....tctttgggttttttcata 3105

ring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

arched: 2552756 seqs, 1349719017 residues
al number of hits satisfying chosen parameters: 5105512

imum DB seq length: 0
imum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : N_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
- 25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3105	100.0	3105	22	Arabidopsis thalia
2	2732.8	88.0	2736	24	Arabidopsis thalia
3	2732.8	88.0	2736	25	Arabidopsis thalia
4	2535.8	81.7	6375	22	Arabidopsis thalia
5	1638.4	52.8	3049	22	Glycine max heat s
6	1564.4	50.4	3052	22	Nicotiana glauca
7	1345	43.3	3084	22	Zea mays heat choc
8	1334.2	43.0	3058	22	Triticum aestivum

Triticum aestivum
Zea mays 101 kDa h
Zea mays heat choc
Phaseolus lunatus
Triticum aestivum
Enterococcus faec
Enterococcus faec
Listeria innocua c
Listeria innocua D
Synchococcus CLPB
Synchococcus sp h
Genomic sequence o
Listeria monocytog
Lactococcus lactis
Haemophilus influ
Haemophilus influ
Staphylococcus aur
Staphylococcus aur
Trypanosoma brucei
High level promote
Leishmania major h
Staphylococcus aur
E. coli DNA for ce
E. coli clpB gene.
Leishmania donovan
Genomic fragment #
Salmonella typhi D
Staphylococcus epi
C glutamicum codin
Corynebacterium gl
C glutamicum codin
S. epidermidis gen
C. trachomatis gen
Complete genome se
C. pneumoniae DNA
Nucleotide sequenc
M. capsulatus gene

ALIGNMENTS

RESULT 1

AAI66065
ID AAI66065 standard; DNA; 3105 BP.

XX AAI66065;

XX 14-JAN-2002 (first entry)

DT Arabidopsis thaliana heat shock protein 101 gene (GenBank: U13949).

DE Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;

KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;

KW Arabidopsis thaliana; ds.

XX Arabidopsis thaliana.

XX WO200170929-A2.

XX 27-SEP-2001.

XX 20-MAR-2001; 2001WO-US08836.

XX 20-MAR-2000; 2000US-190769P.

XX 18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

PI Lindquist S, Queitsch C, Vierling B;

XX WPI; 2001-639123/73.

DR P-PSDB; AAM51659.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 4; Page -: 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAM51651-AAM51671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 3105 BP; 891 A; 541 C; 852 G; 821 T; 0 other;

Query Match	100.08;	Score 3105;	DB 22;	Length 3105;	
Best Local Similarity	100.08;	Pred. No. 0;			
Matches 3105;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
1	AAAGTTATCAATTTTACAACTTACCGGTATAATCTGCTGTTGATTTCTGCAAAAAGAGAA	60			
1	AAAGTTATCAATTTTACAACTTACCGGTATAATCTGCTGTTGATTTCTGCAAAAAGAGAA	60			
61	GACTTTTACCGAAGAGAGTCTCTGCTCATTCAGAGAACTCAACGAAACAAACCCAG	120			
61	GACTTTTACCGAAGAGAGTCTCTGCTCATTCAGAGAACTCAACGAAACAAACCCAG	120			
121	TTCTCATATATCTTTTAAAGGATTACAAAGCTAATCGAAGATGAATCCAGAGAAATTC	180			
121	TTCTCATATATCTTTTAAAGGATTACAAAGCTAATCGAAGATGAATCCAGAGAAATTC	180			
181	ACACACAGACAAACGAGACAAATTCCTACAGCTCAGTGTAGCTGTGAATCGAGGACAT	240			
181	ACACACAGACAAACGAGACAAATTCCTACAGCTCAGTGTAGCTGTGAATCGAGGACAT	240			
241	GCTCAATTCACCTCTTTCGATTTAGCTGTGCTGTTGATCTCTGATCCACCGGTATATTT	300			
241	GCTCAATTCACCTCTTTCGATTTAGCTGTGCTGTTGATCTCTGATCCACCGGTATATTT	300			
301	CCTCAAGCAATCTCTAGTCCCGGTGGCGAGAACGAGCTCAATCTGCTGAAAGAGTGATC	360			
301	CCTCAAGCAATCTCTAGTCCCGGTGGCGAGAACGAGCTCAATCTGCTGAAAGAGTGATC	360			
361	AATCAAGCTTGAAGAGCTTCTTCAATCTCTCAGCTCAGAGTGTGATTTCCAGCGAGT	420			
361	AATCAAGCTTGAAGAGCTTCTTCAATCTCTCAGCTCAGAGTGTGATTTCCAGCGAGT	420			
421	CTAGTCTTATTAAGTGTATCTGCTGCTCAAGCTGTGCTCAGAGTGTGATTTCTACT	480			
421	CTAGTCTTATTAAGTGTATCTGCTGCTCAAGCTGTGCTCAGAGTGTGATTTCTACT	480			
481	CAATTTGGCTGTTGACAGTGTATTTAGGTTCTTTTGAAGATTTCTCAATCAGGATTTG	540			
481	CAATTTGGCTGTTGACAGTGTATTTAGGTTCTTTTGAAGATTTCTCAATCAGGATTTG	540			
541	TTGAAACGAAGTCTGAGGACGCGAGGTTAAAGTCTGAGTTGAGAGCTTCTGTGGG	600			
541	TTGAAACGAAGTCTGAGGACGCGAGGTTAAAGTCTGAGTTGAGAGCTTCTGTGGG	600			
601	AAAGAGGGAGAGAGTCTGAGAGTGTCTCAGGGGACACAAATTTTCAAGCTTTAAAGACT	660			
601	AAAGAGGGAGAGAGTCTGAGAGTGTCTCAGGGGACACAAATTTTCAAGCTTTAAAGACT	660			
661	TATGGAAGAGATTTGGTTGAGCAAGCGAGGAGCTTGTATCTGTGATTTGGTCTGATGAG	720			

661	TATGGAAGAGATTTGGTTGAGCAAGCGAGGAGCTTGTATCTGTGATTTGGTCTGATGAG	720			
721	GAGATTAGAGAGTCTGAGGATTTCTTTCGAGGAGACGAGAGCAAACTCTGTCTTATT	780			
721	GAGATTAGAGAGTCTGAGGATTTCTTTCGAGGAGACGAGAGCAAACTCTGTCTTATT	780			
781	CGAGAGCCAGGAGTTGGTAAAAACAGCTGTGTTGAAGGTTTACGCAAAAGGATTTGAAA	840			
781	CGAGAGCCAGGAGTTGGTAAAAACAGCTGTGTTGAAGGTTTACGCAAAAGGATTTGAAA	840			
841	GAGAGTGTCCCAACAGCTTCTACTGATGTGAGATTAATTTCTGTCGACATGGTGGTTA	900			
841	GAGAGTGTCCCAACAGCTTCTACTGATGTGAGATTAATTTCTGTCGACATGGTGGTTA	900			
901	GTTCGTGGTCTTAAATACCGAGAGAGTTTGAAGAAAGGTTGAAATCTGTTTGAAGAA	960			
901	GTTCGTGGTCTTAAATACCGAGAGAGTTTGAAGAAAGGTTGAAATCTGTTTGAAGAA	960			
961	GTTGAGGACGCTCAAGGCAAAAGTGTATCTCTTTTATTTGATGAGATTCATTTGGTT	1020			
961	GTTGAGGACGCTCAAGGCAAAAGTGTATCTCTTTTATTTGATGAGATTCATTTGGTT	1020			
1021	GCTGGCAAACTGAGAGGTCGATGATGACGCTATCTGTTCAAGCCCATGTTAGCTAGA	1080			
1021	GCTGGCAAACTGAGAGGTCGATGATGACGCTATCTGTTCAAGCCCATGTTAGCTAGA	1080			
1081	GGCAGCTTCGATGCTATCTGCTACACGCTTGAAGAAATACAGGAAATATCTTGAGAA	1140			
1081	GGCAGCTTCGATGCTATCTGCTACACGCTTGAAGAAATACAGGAAATATCTTGAGAA	1140			
1141	GATGCTGCTTTGAGAGGAGTTTCCAAAGTCTATTTGTCGAGGACCAAGTGTGCTGAC	1200			
1141	GATGCTGCTTTGAGAGGAGTTTCCAAAGTCTATTTGTCGAGGACCAAGTGTGCTGAC	1200			
1201	ACCATTAGTATCTTACAGGACTCAAGGAGAGTATGAGGACATCATGCTGTGCGAATC	1260			
1201	ACCATTAGTATCTTACAGGACTCAAGGAGAGTATGAGGACATCATGCTGTGCGAATC	1260			
1261	CAAGACAGAGCTCTTATAAATGCTGCTCAGCTGTCTGCTGCTTACATAACTGCTGGCAT	1320			
1261	CAAGACAGAGCTCTTATAAATGCTGCTCAGCTGTCTGCTGCTTACATAACTGCTGGCAT	1320			
1321	TTACCGGATTAAGCAATTTGATTTGGTTGATGAGGCTTGTGCGAATGTGAGATTCAGCTT	1380			
1321	TTACCGGATTAAGCAATTTGATTTGGTTGATGAGGCTTGTGCGAATGTGAGATTCAGCTT	1380			
1381	GATAGTCAACCTGAGAGATTTGATTAACCTTGAAGGAGAGGATGAGCTGGAATTTGAA	1440			
1381	GATAGTCAACCTGAGAGATTTGATTAACCTTGAAGGAGAGGATGAGCTGGAATTTGAA	1440			
1441	CTTCAAGCTTTGAAAGGAGAGGATTAAGCCAGCAAAAGCTCGACTTATAGAGTGGCG	1500			
1441	CTTCAAGCTTTGAAAGGAGAGGATTAAGCCAGCAAAAGCTCGACTTATAGAGTGGCG	1500			
1501	AAAGAGCTTCAATGACCTGAGAGCAAGCTTTCAGCTCTCAGATGAAATACAGAAAGGAG	1560			
1501	AAAGAGCTTCAATGACCTGAGAGCAAGCTTTCAGCTCTCAGATGAAATACAGAAAGGAG	1560			
1561	AAAGAGAGATTTGATGAGATTTCCAGGCTTAAACAGAAAGAGAGCTCATGTTTCT	1620			
1561	AAAGAGAGATTTGATGAGATTTCCAGGCTTAAACAGAAAGAGAGCTCATGTTTCT	1620			
1621	TTGAGGAGGACGAGACGAGATATGACCTTTCAGAGAGCTGCTGATCTAAGATATGCGCA	1680			
1621	TTGAGGAGGACGAGACGAGATATGACCTTTCAGAGAGCTGCTGATCTAAGATATGCGCA	1680			
1681	ATTCAAGAGTGGAAATCTGCAATTTGCCAACTTTGAGAGAACTTCTTCTGAGAGATG	1740			
1681	ATTCAAGAGTGGAAATCTGCAATTTGCCAACTTTGAGAGAACTTCTTCTGAGAGATG	1740			
1741	ATGCTCAAGAGAGTGGGCTTGACACATTTGAGGCTTGTGAGCTTGTGAGCCCTTGACAGG	1800			

2881	QY	GAATGATCGAGGATTTAAAGTCTTGACCTCCAAAGTGTGAGTCTCTTTTGAGGTTGTGGA	2940
2882	QY		
2883	QY		
2884	QY		
2885	QY		
2886	QY		
2887	QY		
2888	QY		
2889	QY		
2890	QY		
2891	QY	TTGTATGACTTGAGCTTTTCATGCTCGTTTTTGTTTTGACTCTCTTTTGGTGTAGACTATAATG	3000
2892	QY		
2893	QY		
2894	QY		
2895	QY		
2896	QY		
2897	QY		
2898	QY		
2899	QY		
2900	QY		
2901	QY	GTGTGGAACCTGTAAAAAGTCTACTTTATTATGTTTTTCGTATATGATAAATCATTTTACAT	3060
2902	QY		
2903	QY		
2904	QY		
2905	QY		
2906	QY		
2907	QY		
2908	QY		
2909	QY		
2910	QY		
2911	QY		
2912	QY		
2913	QY		
2914	QY		
2915	QY		
2916	QY		
2917	QY		
2918	QY		
2919	QY		
2920	QY		
2921	QY		
2922	QY		
2923	QY		
2924	QY		
2925	QY		
2926	QY		
2927	QY		
2928	QY		
2929	QY		
2930	QY		
2931	QY		
2932	QY		
2933	QY		
2934	QY		
2935	QY		
2936	QY		
2937	QY		
2938	QY		
2939	QY		
2940	QY		
2941	QY		
2942	QY		
2943	QY		
2944	QY		
2945	QY		
2946	QY		
2947	QY		
2948	QY		
2949	QY		
2950	QY		
2951	QY		
2952	QY		
2953	QY		
2954	QY		
2955	QY		
2956	QY		
2957	QY		
2958	QY		
2959	QY		
2960	QY		
2961	QY	AAGAAATCCATTTGATCGTCCCAATCTTTTGGGTTTTTTTTCATA	3105
2962	QY		
2963	QY		
2964	QY		
2965	QY		
2966	QY		
2967	QY		
2968	QY		
2969	QY		
2970	QY		
2971	QY		
2972	QY		
2973	QY		
2974	QY		
2975	QY		
2976	QY		
2977	QY		
2978	QY		
2979	QY		
2980	QY		
2981	QY		
2982	QY		
2983	QY		
2984	QY		
2985	QY		
2986	QY		
2987	QY		
2988	QY		
2989	QY		
2990	QY		
2991	QY		
2992	QY		
2993	QY		
2994	QY		
2995	QY		
2996	QY		
2997	QY		
2998	QY		
2999	QY		
3000	QY		

ABZ12430	ID	ABZ12430 standard; DNA; 2736 BP.
XX	AC	ABZ12430;
XX	DT	21-JAN-2003 (first entry)
XX	DE	Arabidopsis thaliana stress regulated gene SEQ ID NO 235.
XX	KW	Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX	OS	Arabidopsis thaliana.
XX	PN	WO200216655-A2.
XX	PD	28-FEB-2002.
XX	PF	24-AUG-2001; 2001WO-US26685.
XX	PR	24-AUG-2000; 2000US-227866P.
XX	PR	26-JAN-2001; 2001US-26467P.
XX	PR	22-JUN-2001; 2001US-30011P.
XX	PA	(SCTI) SCRIPPS RES INST.
XX	PA	(SYGN) SYNGENTA PARTICIPATIONS AG.
XX	PI	Harper JF, Kreps J, Wang X, Zhu T;
XX	PI	WPI; 2002-304127/34.
XX	PT	Identifying a stress condition to which a plant cell has been exposed
XX	PT	and producing plants with increased tolerance to these abiotic stresses
XX	PS	Claim 144; SEQ ID NO 235; 577pp + Sequence Listing; English.
XX	CC	The invention relates to identifying a stress condition to which a plant
XX	CC	cell has been exposed, comprising:
XX	CC	(a) contacting nucleic acid representative of expressed polynucleotides
XX	CC	in the plant cell with an array or probes representative of the plant
XX	CC	cell genome; and
XX	CC	(b) detecting a profile of expressed polynucleotides in the plant cell
XX	CC	characteristic of a stress response. The method is useful in the
XX	CC	production of transgenic plants, cells and seeds and in producing plants
XX	CC	with increased tolerance to abiotic stress. The present sequence is that
XX	CC	of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
XX	CC	in methods of the invention.
XX	CC	Note: The sequence data for this patent is not represented in the printed
XX	CC	specification but is based on sequence information supplied to Derwent by
XX	CC	the European Patent Office.
XX	SQ	Sequence 2736 BP; 784 A; 481 C; 785 G; 686 T; 0 other;

Query Match 85.0%; Score 2732.8; DB 24; Length 2736;

Best Local Similarity	99.9%; Pred. No. 0;	Matches	2;	Indels	0;	Gaps	0;
atches 2734; Conservative	0;	Mismatches					
163	ATGAATCAGAGAAATTCACACACAGAACAGAGCAATTCGTACAGCTCATGAGCTA	222					
1	ATGAATCAGAGAAATTCACACACAGAACAGAGCAATTCGTACAGCTCATGAGCTA	60					
223	GCTGTGAATCAGAGCAATGCTCAATTCATCTCTTTGCAATTTAGCTGTGCTTTGATCTCT	282					
61	GCTGTGAATCAGAGCAATGCTCAATTCATCTCTTTGCAATTTAGCTGTGCTTTGATCTCT	120					
283	GATCCCAACGGGTATATTTCTCAAGCAATCTCTAGTGCCTGGGCGAGAACGCGAGCTCAA	342					
121	GATCCCAACGGGTATATTTCTCAAGCAATCTCTAGTGCCTGGGCGAGAACGCGAGCTCAA	180					
343	TCCTGCTGAAAGAGTGATCAATCAAGCCCTTGGAAGAGTTCCTTCACAATCTCTCCACCT	402					
181	TCCTGCTGAAAGAGTGATCAATCAAGCCCTTGGAAGAGTTCCTTCACAATCTCTCCACCT	240					
403	GATGATATTCAGAGAGTTCAGTCTTATTAAGGTCAATTCCTGCTCAAGCTGCTCAG	462					
241	GATGATATTCAGAGAGTTCAGTCTTATTAAGGTCAATTCCTGCTCAAGCTGCTCAG	300					
463	AGTCAAGAGGTGATCAATTCCTGCTGTGACCAAGTTCATTTATGGGTCTCTTGAAGAT	522					
301	AGTCAAGAGGTGATCAATTCCTGCTGTGACCAAGTTCATTTATGGGTCTCTTGAAGAT	360					
523	TCCTAAATCAGGAGTTTGTGAACGAAGTCGGTGTAGCGACGGCGAGGGTAAAGTCTGAG	582					
361	TCCTAAATCAGGAGTTTGTGAACGAAGTCGGTGTAGCGACGGCGAGGGTAAAGTCTGAG	420					
583	GTTGAGAGCTTCGTGGGAAAGAGGAGAGAGTTCAGAGTTCCTCAGGGGACACAAAT	642					
421	GTTGAGAGCTTCGTGGGAAAGAGGAGAGAGTTCAGAGTTCCTCAGGGGACACAAAT	480					
643	TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGGAAGCTTGATCCT	702					
481	TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGGAAGCTTGATCCT	540					
703	GTTGATGCTGCTGATCAGAGAGATTAAGAGTTCGTGAGGATTTCTTTCAGGAGAACGAAG	762					
541	GTTGATGCTGCTGATCAGAGAGATTAAGAGTTCGTGAGGATTTCTTTCAGGAGAACGAAG	600					
763	ACAAATCCCTGCTTATTCGAGAGCCAGAGTTCGTAAACAGCTGTGCTTCAAGGTTTA	822					
601	ACAAATCCCTGCTTATTCGAGAGCCAGAGTTCGTAAACAGCTGTGCTTCAAGGTTTA	660					
823	GCACAAAGGATTTGGAAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTCG	882					
661	GCACAAAGGATTTGGAAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTCG	720					
883	TTGGAACATGGGTGCTGATGCTGCTGCTTAATACCGAGGAGAGTTCGAAGAAAGTTG	942					
721	TTGGAACATGGGTGCTGATGCTGCTGCTTAATACCGAGGAGAGTTCGAAGAAAGTTG	780					
943	AAATCTGTTTGAAGAGTTCAGGAGCTGGAAGCAAGTTCCTTTATTCATGAG	1002					
781	AAATCTGTTTGAAGAGTTCAGGAGCTGGAAGCAAGTTCCTTTATTCATGAG	840					
1003	ATTCAATTTGCTTCTGCTGCTGCAAACTGAAGGTCATGAGATCAGCTAATCTGTTTC	1062					
841	ATTCAATTTGCTTCTGCTGCTGCAAACTGAAGGTCATGAGATCAGCTAATCTGTTTC	900					
1063	AGCCCAATGTTAGTAGAGGCGAGCTTCGATGCTGCTGCTTACAGCTTCGAGAAATAC	1122					
901	AGCCCAATGTTAGTAGAGGCGAGCTTCGATGCTGCTGCTTACAGCTTCGAGAAATAC	960					
1123	AGGAATATGTTGAGAAAGTGTGCTTTCAGAGGAGGTTCCAAACAGTCTATGTTGG	1182					
961	AGGAATATGTTGAGAAAGTGTGCTTTCAGAGGAGGTTCCAAACAGTCTATGTTGG	1020					
1183	GAGCCAAAGTGTGCTGACACATTTAGTATCTTCTAGAGGACTCAAGGAGAGTATGAGGA	1242					
1021	GAGCCAAAGTGTGCTGACACATTTAGTATCTTCTAGAGGACTCAAGGAGAGTATGAGGA	1080					
1243	CATCATGTTGTCGGATCCAGACAGAGCTCTTATAAATGCTGCTCAGCTGCTGCTGCT	1302					
1081	CATCATGTTGTCGGATCCAGACAGAGCTCTTATAAATGCTGCTGCTGCTGCTGCTGCT	1140					
1303	TACATAAATGTTGTCGGATCCAGATTAAGCAATTTGATTTGTTGATGAGGCTTTGTCG	1362					
1141	TACATAAATGTTGTCGGATCCAGATTAAGCAATTTGATTTGTTGATGAGGCTTTGTCG	1200					
1363	AATGTGAGAGTCCAGCTTGTATAGTCAACTGAAGAGATTTGATAAATGCTTGAAGGAGAGG	1422					
1201	AATGTGAGAGTCCAGCTTGTATAGTCAACTGAAGAGATTTGATAAATGCTTGAAGGAGAGG	1260					
1423	ATGCACCTGGAATTTCAACTTCACGCTTTGGAAGGAGAGATTAAGCAATGCTTGAAGGAGAGG	1482					
1261	ATGCACCTGGAATTTGAACTTCACGCTTTGGAAGGAGAGATTAAGCAATGCTTGAAGGAGAGG	1320					
1483	CGACTTATAGAGTTCGGAAGAGCTTTGATGACCTGAGAGACAAAGCTTCAGCTCTCAG	1542					
1321	CGACTTATAGAGTTCGGAAGAGCTTTGATGACCTGAGAGACAAAGCTTCAGCTCTCAG	1380					
1543	ATGAATAACAGAAAGAGAGAGAGAGATTTGATGAGATTCGAGGCTTAAACAGAAAGAG	1602					
1381	ATGAATAACAGAAAGAGAGAGAGAGATTTGATGAGATTCGAGGCTTAAACAGAAAGAG	1440					
1603	GAAGAGTCTCATGTTTCTTTGTCAGGAGGAGAGAGAGATTAAGCAATGCTTGAAGGAGCT	1662					
1441	GAAGAGTCTCATGTTTCTTTGTCAGGAGGAGAGAGAGATTAAGCAATGCTTGAAGGAGCT	1500					
1663	GATCTAAGATATTCGCGCAATTCAGAAAGTGAATCTGCAATTCGCCCAACTTGAAGGAGCT	1722					
1501	GATCTAAGATATTCGCGCAATTCAGAAAGTGAATCTGCAATTCGCCCAACTTGAAGGAGCT	1560					
1723	TCCTTCTGAAGAGATGCTGATGCTCAGAGAAAGCTTGGGCTCGAACACATTTGCTGAGGTT	1782					
1561	TCCTTCTGAAGAGATGCTGATGCTCAGAGAAAGCTTGGGCTCGAACACATTTGCTGAGGTT	1620					
1783	GTGAGCCCTTGGACAGGAGATTCAGAGTTCAGAGATTCGCGCAAAATGAGAGGAGAGGTTG	1842					
1621	GTGAGCCCTTGGACAGGAGATTCAGAGTTCAGAGATTCGCGCAAAATGAGAGGAGAGGTTG	1680					
1843	ATTGGTCTTCTGATAGTTCGATAGCGGGTTGCGGACAGAGATCAAGCGGTAAATGCA	1902					
1681	ATTGGTCTTCTGATAGTTCGATAGCGGGTTGCGGACAGAGATCAAGCGGTAAATGCA	1740					
1903	GTTTCTGAGGCAATTCCTAAGGTCAAGGCGAGAGCTTGGTAGGCAACACAGCCAACTGGA	1962					
1741	GTTTCTGAGGCAATTCCTAAGGTCAAGGCGAGAGCTTGGTAGGCAACACAGCCAACTGGA	1800					
1963	TCATTTCTTATTTCTTGGACCAACTGTTGTCGCAAACTCAGCTCGCAAGGCTCTTGCT	2022					
1801	TCATTTCTTATTTCTTGGACCAACTGTTGTCGCAAACTCAGCTCGCAAGGCTCTTGCT	1860					
2023	GAGCAGCTGTTTGTATGATGAAACCTCTTATGTTGCGATTCATATGTCGGAATATATGGA	2082					
1861	GAGCAGCTGTTTGTATGATGAAACCTCTTATGTTGCGATTCATATGTCGGAATATATGGA	1920					
2083	CAACACTGCTCTCTCGCTCATTTGGGCGCACCAAGGCTATGTTGCTCACGAGGAGGT	2142					
1921	CAACACTGCTCTCTCGCTCATTTGGGCGCACCAAGGCTATGTTGCTCACGAGGAGGT	1980					
2143	GGACACTAATCTGAGGCTGTGAGGAGCGACCTTATGTTGTCATATCTTTTGTATGAGTG	2202					
1981	GGACACTAATCTGAGGCTGTGAGGAGCGACCTTATGTTGTCATATCTTTTGTATGAGTG	2040					
2203	GAGAGGCTCATGTTGCTGCTTCAACTCTGCTCCAGGTTTGGATGATGCTGCTGATG	2262					
2041	GAGAGGCTCATGTTGCTGCTTCAACTCTGCTCCAGGTTTGGATGATGCTGCTGATG	2100					
2263	ACAGACGGGCAAGGCGAGAGCTCGATTTTCAAGGAACCTCGGTGATAATCATGACATCAAC	2322					
2101	ACAGACGGGCAAGGCGAGAGCTCGATTTTCAAGGAACCTCGGTGATAATCATGACATCAAC	2160					

2323 CTTGCTGCTGACACCTCTCTTGCAGGGCTAACTGGGAAAGTAACAATGGAAGTGGCCCGG 2382
 2161 CTTGCTGCTGACACCTCTCTTGCAGGGCTAACTGGGAAAGTAACAATGGAAGTGGCCCGG 2220
 2383 GACTGTGTGATGCGGAGGTGAGGAACACTTCAGACAGAGCTCTTGAACAGGCTTGAC 2442
 2221 GACTGTGTGATGCGGAGGTGAGGAACACTTCAGACAGAGCTCTTGAACAGGCTTGAC 2280
 2443 GAGATTGTGTGTTCTGACCCCTTTCACATGACAGTTGAGGAAAGTAGCTCGGCTTCAA 2502
 2281 GAGATTGTGTGTTCTGACCCCTTTCACATGACAGTTGAGGAAAGTAGCTCGGCTTCAA 2340
 2503 ATGAAGACGTTGTCTGCTCGGCTTCTGCTGAAAGAGAGTGTCTTTGGCAGTCACTGATGCT 2562
 2341 ATGAAGACGTTGTCTGCTCGGCTTCTGCTGAAAGAGAGTGTCTTTGGCAGTCACTGATGCT 2400
 2563 GCTTTGGACTATATCTTGGCAGAGTATGACCCGGTGTATGTGTCTAGGCTTAAAG 2622
 2401 GCTTTGGACTATATCTTGGCAGAGTATGACCCGGTGTATGTGTCTAGGCTTAAAG 2460
 2623 AGATGGATGGAGAGAGTGTCTGACAGAACTGTCAAAGATGTTGTGCGTGAGGAAATC 2682
 2451 AGATGGATGGAGAGAGTGTCTGACAGAACTGTCAAAGATGTTGTGCGTGAGGAAATC 2520
 2683 GATGAAACTCCACTGTTTACATAGATGACGGCGCTGTGTATCTTGTATCCGGGTAGAA 2742
 2521 GATGAAACTCCACTGTTTACATAGATGACGGCGCTGTGTATCTTGTATCCGGGTAGAA 2580
 2743 AGTGAGTCTAGTGGAGCTTCAACAGGCAAGAGTCAGATGTGCTGATTCATATTGCT 2802
 2581 AGTGAGTCTAGTGGAGCTTCAACAGGCAAGAGTCAGATGTGCTGATTCATATTGCT 2640
 2803 AACGGCCAAAGAGAGTGTGACGCTCAGGCGGTGAAGAGATGAGGATCGAGGAAATA 2862
 2641 AACGGCCAAAGAGAGTGTGACGCTCAGGCGGTGAAGAGATGAGGATCGAGGAAATA 2700
 2863 GAAGATGACGATAATGAGGAAATGATCGAGGATTAA 2898
 2701 GAAGATGACGATAATGAGGAAATGATCGAGGATTAA 2736

UJT 3

AB242033

AB242033 standard; cDNA; 2736 BP.

AB242033;

27-FEB-2003 (first entry)

Arabidopsis thaliana gene #17 modulated by PTGS.

Posttranscriptional gene silencing; PTGS; plant; transformation; gene;

ss.

Arabidopsis thaliana.

Key Location/Qualifiers

CDS 1..2736

/*tag= a

W0200291695-A2.

17-OCT-2002.

05-APR-2002; 2002WO-EP03806.

06-APR-2001; 2001US-282049P.

(SYGN) SYNGENTA PARTICIPATIONS AG.

(FRIE-) FRIEDRICH MIESCHER INST.

Zhu T, Glazov EA, Meins F, Wang X, Chang H;

XX WPI; 2003-103337/09.
 DR P-PSDB; ABP81189.
 XX Novel polynucleic acid segment useful for modulating gene expression
 PT within a cell by posttranscriptional gene silencing, and for augmenting
 PT a plant cell genome -
 XX Claim 18; Page 185-186; 438pp; English.
 XX The invention relates to a novel isolated polynucleic acid segment
 CC modulated within a cell by posttranscriptional gene silencing (PTGS). The
 CC invention specifically relates to a method to identify an expression
 CC product that is modulated by PTGS. The polynucleotide is useful for
 CC modulating the gene expression within a cell by PTGS, by introducing the
 CC polynucleic acid into a cell and expressing the nucleic acid segment in
 CC the cell to form a product. The polynucleic acid segment is also useful
 CC for augmenting a cell genome, and for augmenting a plant genome, by
 CC contacting a plant cell with the segment to produce a transformed plant
 CC cell, and growing the transformed plant cell to produce a differentiated
 CC transformed plant. The sequences shown in AB242017 - AB242142 represent
 CC segments of A. thaliana cDNA modulated by PTGS.
 XX Sequence 2736 BP; 784 A; 481 C; 785 G; 686 T; 0 other;
 SQ

Query Match 88.0%; Score 2732.6; DB 25; Length 2736;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2734; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 163 ATGAATCCAGAGAAATTCACACAGCAAAACGAGCAAAATGCTACAGCTCATGAGCTA 222
 Db 1 ATGAATCCAGAGAAATTCACACAGCAAAACGAGCAAAATGCTACAGCTCATGAGCTA 60
 QY 223 GCTGTGAATGCGAGGACATGCTCAATTCCTCCCTTGGCAATTTAGTGTGCTTTGATCTCT 282
 Db 61 GCTGTGAATGCGAGGACATGCTCAATTCCTCCCTTGGCAATTTAGTGTGCTTTGATCTCT 120
 QY 283 GATCCCAACCGGTATATTTCTCAAGCAATCTCTAGTGGCGGTGGGAGAACGACGCTCAA 342
 Db 121 GATCCCAACCGGTATATTTCTCAAGCAATCTCTAGTGGCGGTGGGAGAACGACGCTCAA 180
 QY 343 TCTGCTGAAAGAGTGTATCAATCAAGCCTTGAAGAGCTTCTCTCAATCTCTCCACCT 402
 Db 181 TCTGCTGAAAGAGTGTATCAATCAAGCCTTGAAGAGCTTCTCTCAATCTCTCCACCT 240
 QY 403 GATGATATTCACGCGAGTCTAGTCTTATTAAGGTCAATTCGCTGCTCAAGCTGCTCAG 462
 Db 241 GATGATATTCACGCGAGTCTAGTCTTATTAAGGTCAATTCGCTGCTCAAGCTGCTCAG 300
 QY 463 AAGTCAAGAGTGATATCTCAATTTGGCTGTGACCAAGTTGATTATGGGTCTTCTTGAAGAT 522
 Db 301 AAGTCAAGAGTGATATCTCAATTTGGCTGTGACCAAGTTGATTATGGGTCTTCTTGAAGAT 360
 QY 523 TCTCAATCAGGGATTTGTTGAACAGTCGCTGTAGCGACGCGAGGGTAAAGTCTGAG 582
 Db 361 TCTCAATCAGGGATTTGTTGAACAGTCGCTGTAGCGACGCGAGGGTAAAGTCTGAG 420
 QY 583 GTTGAGAAGCTTCGTGGGAAAGAGGGAAGATTGAGAGTGTGAGAGTGTTCAGGGGACACAAAT 642
 Db 421 GTTGAGAAGCTTCGTGGGAAAGAGGGAAGATTGAGAGTGTTCAGGGGACACAAAT 480
 QY 643 TTTCAAGCTTTAAAGACTTATGGAAGAGATTGGTTGGTGAAGCAAGCGGGAAGCTTGATCCT 702
 Db 481 TTTCAAGCTTTAAAGACTTATGGAAGAGATTGGTTGGTGAAGCAAGCGGGAAGCTTGATCCT 540
 QY 703 GTGATTGCTGCTGATGAGGAGATTAGAGAGTCGTGAGGATCTTTTCAGGAGAACGAG 762
 Db 541 GTGATTGCTGCTGATGAGGAGATTAGAGAGTCGTGAGGATCTTTTCAGGAGAACGAG 600
 QY 763 AACAACTCCTGCTTATTTGGAGAGCCGAGGTTGCTTAAACAGAGCTGTGCTTCAAGGTTTA 822
 Db 601 AACAACTCCTGCTTATTTGGAGAGCCGAGGTTGCTTAAACAGAGCTGTGCTTCAAGGTTTA 660

14-JAN-2002 (first entry)

Arabidopsis thaliana heat shock protein 101 gene (GenBank: AF218796).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton; canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato; *Arabidopsis thaliana*; ds.

Arabidopsis thaliana.

WC200170929-A2.

27-SEP-2001

20-MAR-2001: 2001WO-US08836.

20-MAR-2000: 2000US-190769P.

18-APR-2000: 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI: 2001-639123/73.

P-PSDB: AAM51659.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 4: Page -: 91pp: English.

The invention relates to a transgenic plant, comprising a genetic construct comprising operatively linked to a nucleic acid sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAMS1651-AAMS1671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or *Arabidopsis thaliana*. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 6376 BP: 1914 A: 1150 C: 1470 G: 1842 T: 0 other:

81.7%: Score 2535.8: DB 22: Length 6376:

Local Similarity 85.7%; Pred No. 0;

Matches 3103; Conservative 0; Mismatches 2; Indels 516; Gaps 5;

1. The first group of respondents (n = 10) was composed of students who had completed the course and were currently employed in a related field. 2. The second group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 3. The third group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 4. The fourth group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 5. The fifth group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 6. The sixth group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 7. The seventh group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 8. The eighth group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 9. The ninth group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 10. The tenth group (n = 10) was composed of students who had completed the course and were currently employed in a related field.

1 AAAGTTATCAATTTACAACATTTACCGCTATAATCTGCTTGATTCTCTGC AAAAGAGAA 60

[illegible]

806 AAAGTATACAAATTTACCAACATTAACCGCTATAATCTGCTTGAATCTCTGCAATAAAGAGATA 923

61 GACTTTTACCGAGAAGAGTCCTCTGGCTCATTTGAAGAAACTCAACGAAACAAACCCAG 120

926 GACTTTTACCGAGAAGTCTCTGGCTCATTTGAAGAACTCAACGAACAACCCAG 985

131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

[illegible]

986 TTCTCATATATCGTTTAAAGGTAAATGATCGCGACAACTCTGTTCCTCATTTGTGTGTTT 1045

-----138

1046 TGTGTTGTGATTAGGGTTTACAAAGATACTGAGATTAGTTTTTTTTTTTTTTT 1105

.....

139 --AGGGATTACAAAGCTAATCGAAGATGAATCCAGAGAAATTCACACACAAGACAAACG 196

[illegible]

II06 T CAGGGA T TACAAAGC TAA TCGAAGAT GAA TCCAGAGAA A T T CAGACACAAGACCA AACG II05

Qy	197	AGACAAATTCGTCACAGCTCATGAGCTAGCTGTGAATCGACGACATGCTCAATTCACCTCCTT	256
Db	1166	AGACAAATTCGTCACAGCTCATGAGCTAGCTGTGAATCGACGACATGCTCAATTCACCTCCTT	1225
Qy	257	TGCATTTAGCTGTGCTTTTGATCTCTGATCCCAACCGGTATATTTTCCTCAAGCAATCTCTA	316
Db	1226	TGCATTTAGCTGTGCTTTTGATCTCTGATCCCAACCGGTATATTTTCCTCAAGCAATCTCTA	1285
Qy	317	GTCCCGGTGGCGAGAACGCAAGCTCAATCTGTCTGAAAGAGTGATCAATCAAGCCCTTGAGA	376
Db	1286	GTCCCGGTGGCGAGAACGCAAGCTCAATCTGTCTGAAAGAGTGATCAATCAAGCCCTTGAGA	1345
Qy	377	AGCTTCTCTTCCACAAATCTCTCTCCACCTTGATGATATCCACGCGAGTCTCTAGTCTTTATAAGG	436
Db	1346	AGCTTCTCTTCCACAAATCTCTCTCCACCTTGATGATATCCACGCGAGTCTCTAGTCTTTATAAGG	1405
Qy	437	TCATTCGTCTGTCTCAAGCTGTCTCAGAAGTCAAGAGTGATCTCAATTTGGCTGTTGAC	496
Db	1406	TCATTCGTCTGTCTCAAGCTGTCTCAGAAGTCAAGAGTGATCTCAATTTGGCTGTTGAC	1465
Qy	497	AGTTGATTATGGGTCTTCTTTGAAGATTTCTCAAAATCAGGGATTTTGTGAACGAAGTCGGTG	556
Db	1466	AGTTGATTATGGGTCTTCTTTGAAGATTTCTCAAAATCAGGGATTTTGTGAACGAAGTCGGTG	1525
Qy	557	TAGCGACGGCGGGTAAAGTCTGAGGTTGAGAAGCTTCGTGGGAAAGAGGGAAGAAAG	616
Db	1526	TAGCGACGGCGGGTAAAGTCTGAGGTTGAGAAGCTTCGTGGGAAAGAGGGAAGAAAG	1585
Qy	617	TTGAGAGTGCTTCAGGGGACACAAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTTGG	676
Db	1586	TTGAGAGTGCTTCAGGGGACACAAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTTGG	1645
Qy	677	TTGAGCAACAGAGGAGAGCTTGATCTCTGATTTGGTCGTGATGAGGAGATTTAGAAGAGTCG	736
Db	1646	TTGAGCAACAGAGGAGAGCTTGATCTCTGATTTGGTCGTGATGAGGAGATTTAGAAGAGTCG	1705
Qy	737	TGAGGATTCCTTTTCGAGGAGAACGAAGACAAATCCTCTGTCTTATTTGAGAGCCAGGAGTTG	796
Db	1706	TGAGGATTCCTTTTCGAGGAGAACGAAGACAAATCCTCTGTCTTATTTGAGAGCCAGGAGTTG	1765
Qy	797	GTAATAACAGCTGTGTTGAAGGTTTAGCAAAAGGATTTGAAAGAGATGTGCCCAACA	856
Db	1766	GTAATAACAGCTGTGTTGAAGGTTTAGCAAAAGGATTTGAAAGAGATGTGCCCAACA	1825
Qy	857	GTCTTACTGATGTGAGATTAAATTCGTGGACATCGGTGCTTAGTTGCTGTGCTCTAAAT	916
Db	1826	GTCTTACTGATGTGAGATTAAATTCGTGGACATCGGTGCTTAGTTGCTGTGCTCTAAAT	1885
Qy	917	ACCGAGAGAGTTTGGAAGAAAGTTGAAATCTGTTTGTGAAAGAGTTGAGGACGCTGAAG	976
Db	1886	ACCGAGAGAGTTTGGAAGAAAGTTGAAATCTGTTTGTGAAAGAGTTGAGGACGCTGAAG	1945
Qy	977	GCAAGTGATTCCTTTTATTGATGAGATTCATTTGTTGCTCTGTGCTGCGCAAACTGAAG	1036
Db	1946	GCAAGTGATTCCTTTTATTGATGAGATTCATTTGTTGCTCTGTGCTGCGCAAACTGAAG	2005
Qy	1037	GGTCGATGGAATGACGCTAAATCTGTCTCAAGCCCAATGTTAGCTAGAGGGCAGCTTCGATGCA	1096
Db	2006	GGTCGATGGAATGACGCTAAATCTGTCTCAAGCCCAATGTTAGCTAGAGGGCAGCTTCGATGCA	2065
Qy	1097	TTGGTGCTCAACGCTTTGAAGAAATACAGGAAATATGTTGAGAAAGATGCTGCTTTTGAGA	1156
Db	2066	TTGGTGCTCAACGCTTTGAAGAAATACAGGAAATATGTTGAGAAAGATGCTGCTTTTGAGA	2125
Qy	1157	GGAGGTTCCAAACAGTCTATGTTTGGGAGCCAAAGTGTGCTGCACACATTAGTATCTCTTA	1216
Db	2126	GGAGGTTCCAAACAGTCTATGTTTGGGAGCCAAAGTGTGCTGCACACATTAGTATCTCTTA	2185
Qy	1217	GAGGACTCAAGGAGAGATGAGGGAATCATCATGGTGTGCGAATCCAAAGACAGAGCTCTTA	1276
Db	2186	GAGGACTCAAGGAGAGATGAGGGAATCATCATGGTGTGCGAATCCAAAGACAGAGCTCTTA	2245

[illegible]

3326	TCCTTTTAGTTTCAANAATCTTCTGATTATGAAAAATATTTTCTTCTTACTTACACGATTA	3388
2120	-----GGTATCTTGGTCAACGAGGAAGGTGGACAACTAACTCAGCGCTGTGAGGAG	2168
3386	ACACTTTCTCAGGTATGTGGTCAACGAGGAAGGTGGACAACTAACTCAGCGCTGTGAGGAG	3445
2169	GGGACCTTATTTGGTCAATCTCTTTGATGAGAGTGGAGAGGCTCATGTTGCTGTCTTCAA	2228
3446	GGGACCTTATTTGGTCAATCTCTTTGATGAGTGGAGAGGCTCATGTTGCTGTCTTCAA	3505
2229	CACCTGCTCCAAAGTTTTCGATGATCGTTCGATTGACAGACGGGCAAGCAGGACAGTCCGA	2288
3506	CACCTGCTCCAAAGTTTTCGATGATCGTTCGATTGACAGACGGGCAAGCAGGACAGTCCGA	3565
2289	TTTCAGGAATCCGGTGATTAATCATGACATCAAACTTCGGTGCTGAAACACTCTCTTCGAGG	2348
3566	TTTCAGGAATCCGGTGATTAATCATGACATCAAACTTCGGTGCTGAAACACTCTCTTCGAGG	3625
2349	GCTTACTCGGAAAGTAACATGAAGTGGCGCCGGGACATGCTGTGATCGGGAGGTGAGGA	2408
3626	GCTTACTCGGAAAGTAACATGAAGTGGCGCCGGGACATGCTGTGATCGGGAGGTGAGGA	3685
2409	ACACTTCAGACACAGAGCTCTTGAAACAGGCTTCACAGAGATTTGTGGTGTTCACACCCCTTC	2468
3686	ACACTTCAGACACAGAGCTCTTGAAACAGGCTTCACAGAGATTTGTGGTGTTCACACCCCTTC	3745
2469	ACATGACCAAGTTGAGGAAAGTACTCGGCTTCAAATGAAAGACGTTGCTGTCGGCTTGC	2528
3746	ACATGACCAAGTTGAGGAAAGTACTCGGCTTCAAATGAAAGACGTTGCTGTCGGCTTGC	3805
2529	TCGAAGAGGAGTCTCTTTGGACGATCACTCATGCTGCTTTGGACATATATCTTTGGCAGAG	2588
3806	TCGAAGAGGAGTCTCTTTGGACGATCACTCATGCTGCTTTGGACATATATCTTTGGCAGAG	3865
2589	TTATGACCC-----	2597
3866	TTATGACCCGGTAACTCTTTCACATGATGCGAATGCATATCATTTGTTGATCTCTCAGG	3925
2598	-----GGTGTAT	2604
3926	CTGGTTTCTTCTTGTGACTGACTCTTTGTGGTGTGCTTTCTTTGTTGTTGTTAGGTGTAT	3985
2605	GGTGCTAGCCCTATAAGGACATCGATGAGAGAAAGGTGGTGAACAGAACTGTCAAAGATG	2664
3986	GGTGCTAGCCCTATAAGGACATCGATGAGAGAAAGGTGGTGAACAGAACTGTCAAAGATG	4045
2665	GTGTGCGTGAGGAAATCGATGAAATCTCCACTGTTTACATAGATGACAGCGCTGATGAT	2724
4046	GTGTGCGTGAGGAAATCGATGAAATCTCCACTGTTTACATAGATGACAGCGCTGATGAT	4105
2725	CTTTGTACCGGTAGAAAGTGGAGCTCTAGTGGAGCTTCAAACAGGCAAGAGTCAAGAT	2784
4106	CTTTGTACCGGTAGAAAGTGGAGCTCTAGTGGAGCTTCAAACAGGCAAGAGTCAAGAT	4165
2785	GTGCTGATTCATATGCTAAACGGGCCAAAGAAAGTGAATGACAGCTCAGGCGGTCAAGAG	2844
4166	GTGCTGATTCATATGCTAAACGGGCCAAAGAAAGTGAATGACAGCTCAGGCGGTCAAGAG	4225
2845	ATGAGGATCGAGGAAATGAGAGATGACGATAATGAGGAAATGATCGAGGATTTAAGTCTT	2904
4226	ATGAGGATCGAGGAAATGAGAGATGACGATAATGAGGAAATGATCGAGGATTTAAGTCTT	4285
2905	GACCTCCAAAGTGTGAGTCTTTTGGAGTGTGTCATTTGATGATGACTTGAAGCTTCATGCT	2964
4286	GACCTCCAAAGTGTGAGTCTTTTGGAGTGTGTCATTTGATGATGACTTGAAGCTTCATGCT	4345
2965	CTGTTTTGGTTTTGACTCTTTTGGTGAAGTAAATGGTGTGGAACCTGTAAAAAGTCTAC	3024
4346	CTGTTTTGGTTTTGACTCTTTTGGTGAAGTAAATGGTGTGGAACCTGTAAAAAGTCTAC	4405
3025	TTTATATATGTTTTCGTATATGTAATAATCAATTTTACATAAGAAATCCATTTGATGCTCCATA	3084

4406 TTATTATGTTTCGTATATGTAATCAATTTATACATAAGAAATCCATTTAGTACGTCCCATTA 4465

3085 ATCTTTTGGGGTTTTTTCATA 3105

4466 ATCTTTTGGGGTTTTTTCATA 4486

ULT 5
6067

AAI66067 standard; DNA; 3049 BP.

AAI66067;

14-JAN-2002 (first entry)

Glycine max heat shock protein gene (GenBank: L35272).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
Arabidopsis thaliana; ds.

Glycine max.

W0200170329-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

P-PSDB; AAM51662.

Transgenic plants with improved heat stress tolerance, useful for
producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic
construct comprising a promoter operatively linked to a nucleic acid
sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
family amino acid sequence (AAM51651-AAI66071). The transgenic plant has
increased stress tolerance, especially to heat. The plant is a cereal,
grass, ornamental plant, crop plant, food plant, oil-producing plant, a
synthetic product-producing plant, an environmental waste absorbing
plant, an alcohol plant, a medicinal plant, a recreational plant and/or
an animal feed plant. In particular, the transgenic plant is cotton,
canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
Arabidopsis thaliana. The plants may be used to produce animal feed,
alcohol, crop, oil, medicine or a synthetic product.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained from GenBank using the Accession Number
reference provided in the specification.

Sequence 3049 BP; 850 A; 564 C; 863 G; 772 T; 0 other;

Query Match 52.8%; Score 1638.4; DB 22; Length 3049;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 2094; Conservative 0; Mismatches 646; Indels 18; Gaps 5;

161 AGATGAATCCAGAGAAATTCACACAGACAAAGAGACAAATTTGCTACAGCTCATGAGC 220

134 AAATGAATCTCTGAGAGTTTACTCACAGACAAATGAAGCTCTTTTCGCTCGGCCACGAGC 193

221 TAGCTGTGAATGCAGGACATCTCAATTCATCTCTTTCATTTAGCTGGTGTGATCT 280

194 TCGCGATAGTTAGGCGCACCGCAATTTGACTCCATCCACTTGGCCCCACCGCTGATT 253

281 CTGATCCACCGGTATATTTCTCAAGCAATCTCTAGTGC---CGTGGCGAGAACCGAG 337
254 CGATCCCAACCGCATCTTCGTCTAGCGATAAACAAGCGCGCGCGCGAGGAATCGG 313
338 CTCAATCTGCTGAAAGAGTGATCAATCAAGCCTTGAAGAGCTTCCCTTCAAAATCTCTC 397
314 CACGCGCGGTGAGCGAGTGTGAAACCAAGGCTCTGAAGAAAGTACCTCCCTGCCAGTCCCTC 373
398 CACCTGATGATATTTCCAGCGAGTTCCTAGTCTTATTAAGGTCAATTCGTCTCAAGCTG 457
374 CGCCGACGAGTTCGCGCGAGCACCAACCTCTGTAGGGCCATCAGGAGAGCACAGCGG 433
458 CTCAGAGTCAAGAGGTGATATCTATTTGGTGTGTGACCAAGTGTATTAAGGTCTTCTT 517
434 CGCAAAATCAAGTGGCGACACAGCGGTTCGCGGTGTGATCAGTGTATCTCTCGGAATCTCT 493
518 AAGATCTCAATCAGGAGTTCGTTGACGAAGTCCGTGTAGCGACGCGGAGGTTAAAGT 577
494 AAGACTCCCAATCGGAGACCTGTGAAAGGAAGCGGGTTCGCGTGGCGAAGGTAGAA 553
578 CTGAGTTGAGAAGCTTCGTGGGAAAGAGGGAAGAGTTTGAGAGTCTTTCAGGGGACA 637
554 CGGAAGTGTATAGCTTCGTGGGAGGAAGGAGGAGTTGAGAGCGCTTCCGGGGATA 613
638 CAAATTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGAGAGCTTG 697
614 CGAATTTCAAGCTTTGAAGACTTATGGCGGTGACCTTGTGTTGAAACAGCGGGGAGCTCG 673
698 ATCCGTGATTTGGTCTGATGAGGAGATTTAGAGAGTCTGTCAGGATTTCTTCGAGGAGAA 757
674 ACCCTGTTATGGCGGTGACGAAGAGATTTAGAGGTTGTGAGGATTTCTATCAGCGGGA 733
758 CGAAGACAATCTCTGTCTTATTTGAGAGCCAGGAGTTGGTTAAACAGCTGTGGTTGAAG 817
734 CTAAGACAACCCGGTCTCTGTTGAGAACCCGGTGTGGGAAACCTGCGGTGTGGGAG 793
818 GTTTAGCAAAAGGATTTGAAAGAGATGTGCCCAACAGTCTTATCTGATGTGAGATTAA 877
794 GGTTCGACAGAGGATAGTAAGAGGGGATGTTCACAGCAACCTTGTGATGTGAGGCTTA 853
878 TTTTCGTTGACATGGTGGCTAGTTCGTGCTGTCTTAATATACGAGGAGAGTTTGAAGAA 937
854 TTGCGTTGATATGGGGCGTGTGTCGGGTGTCAGTATAGAGCGAGTTTGAAGAGC 913
938 GGTTCAAATCTGTTTGAAGAGTGTGAGGACCGCTGAAAGGCAAAAGTCTTCTTTATTG 997
914 GGTAAAGCTGTTTGAAGAGTGTGAGGAGGCTGAGGGGAGGATGATCTCTTCAATTG 973
998 ATGAGATTCATTTGTTCTGTTGCTGGCAAAACGAAAGGTCGATGATGAGTAAATC 1057
974 ATGAGATTCATTTGTTCTGTTGCTGGTAGAACTGAAAGGCTCCATGATGATGCTGCTAATC 1033
1058 TCTTCAAGCCCATTTAGTACGAGGCGAGCTTCGATGATTTGGTGTCTACAAAGCTTGAAG 1117
1034 TATTCAAACTATGTTCTGCGGCGAGCTTAGGTGATTTGGTGGCCACACAGCTTGAGG 1093
1118 AATACAGGAAATATGTTGAGAAAGATGCTGCTTTGAGAGGAGGTTTCCAAAGTCTATG 1177
1094 AGTACAGGAGTATGTGGGAGAGGATGCTGCATTCGAGAGGAGGTTCCACAGGTTTGTG 1153
1178 TTGCGGAGCAAGTGTGCTGACACCTATGATCTCTTAGAGGACTCAAGGAGAGATGATG 1237
1154 TGGCGGAACTAGTGTGGTGTATACCATTTAGCATTTCTTCGTGGCTTGAAGAGAGATG 1213
1238 AGGACATCATGCTGTGGAAATCCCAAGACAGAGCTCTTATAAATGCTGCTCAGCTGTCTG 1297
1214 AAGGCCATCACGGTGTGAGAAATTCAGGACCGGTCTTGGTTATGGAGCTCAATTGCTA 1273
1298 CTCGTTTACATACTCGTGGCATTTTACCGGATTAAGCAATTTGATTTGGTTGATGAGGCTT 1357
1274 ACCGGTATATATCTGGCGGTCTATCTCTCTGACAGGCAATGATTTGGTTGACGAGGCTT 1333

2411	Db	TTGATGAAATTGGTATTGACCCCTTTTCATGACCAATTGAGGAGTTGCAGGC	2470
2498	QY	TTCAAAATGAAGACGTTGCTGTCGCGCTTGCTGAAGAGGAGTGTCTTTGGCAGTCACTG	2557
2471	Db	TACAAATGAAGGATGTAGCTAGTCGTCTTGCTGAGAAAGGAATTGCTTTGGCAGTCACTG	2530
2558	QY	ATGCTGCTTTGGACTATATCTTTGGCAGAGAGTTATGACCCGGTGATATGGTGCTAGGCCTA	2617
2531	Db	ATGCAGCACTAGACTATATACCTTTCCGAGAGCTATGATCCGGTGTATGGTGCTAGACCAA	2590
2618	QY	TAAAGAGATGGATGGAGAGAGAGGTGGTGACAGAACTGTCAAGAGATGGTTGTGCGTGAGG	2677
2591	Db	TAAAGAGAGTGGCTTGAGAGAGAGAGTGGTGACAGAGTTGCTCAGAATGCTTGTAAAGAGAGG	2650
2678	QY	AAATCGATGAAACTCCACTGTTTTCATAGATCCAGG-----CGCTGGTGATCTTGTGT	2731
2651	Db	AGATTGATGAGAAATTCAACGGTTTACATGATGCTGGACCCAAATGGGGCGAGTTGGTCT	2710
2732	QY	ACCGGCTAGAAA---AGTGGAGGCTTAGTGACGCTTCAACAGGCAAGAGTCAAGATGTGC	2788
2711	Db	ACCGTGTAGAAAGAATGGAGGGTTGTTAATCTTACACTGGGCAGAGTCCGATATCT	2770
2789	QY	TGATTCAATATGCTAAACGGC-----CAAAGAGAGTGTATGCAGCTCAGCGCGGTCAAGAAGA	2845
2771	Db	TGATTCAGATACCTAATGACATGCACCTAAAACCTGATGCTTCAAGCAGTCAAGAAGA	2830
2846	QY	TGAGGATCGAGGAAATAGAAAGATGACGATTAATGAGGAAATGATCGAGGATTAAGTCT	2903
2831	Db	TGAAGATTGAGGAAATTGATGATGATGAAATGGAAGAGTGAATTTTATGATAAAATGT	2888

AA166068

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AA156057-AA156084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAW51651-AAW51671). The transgenic plant has

increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 3052 BP; 898 A; 536 C; 836 G; 782 T; 0 other;

Query Match 50.4%; Score 1564.4; DB 22; Length 3052;
Best Local Similarity 74.4%; Pred. No. 0;
Matches 2044; Conservative 0; Mismatches 686; Indels 18; Gaps 5;

147 CAAAGCTAATCGAAGATGAATCCAGAGAAATTACACACACAGACAAACGAGCAATTGC 206
85 CTATAGGTGACTATTATGAATCTCGAAATTCACCCACAGACTAACGAGGCCCTTGC 144

207 TACAGCTCATGAGCTAGCTGTGAATGACGAGACATGCTCAATTCTCTTTGCAATTAGC 266
145 TGGGGCACTCGAGCTAGCACTATCCGAGGGGCATGCTCAATTTACGCTCTGCATATGC 204

267 TGGTCTTTCATCTCTCATCCACCGGTATATTCTCAAGCAATCTCTAGTCCCGGTG 326
205 TGTGGCTTAAATCTCATCATATGTTATTTTCGACAGCGATGTCATGCTGGTGG 264

327 ---CGAAGCGAGCTCAATCTGCTGAAGAGTGATCAATCAAGCCTTGAAGAGCTTCC 383
265 GAATGAAGAGTAGTAATTCAGTGGAGCGGTATTGAATCAAGCGATGAAGAAGTACC 324

384 TTCCATCTCTCCACTCATGATATTCAGCGAGTTCAGTCTTATTAAGTCAATTGC 443
325 TTCTCAACACCGGCTCTGACGAAATCCCACTAGCACTTCCCTTATCAAGGTGTACG 384

444 TCGTCTCAAGCTGCTCAGAGTCAAGAGTGATCTCAATTTGGCTGTTGACCACTTGAT 503
385 CCGAGCAATCTGCGCAGAGTCTGCTGTGACAGCCATTTAGCAGTGGATCAGTTGAT 444

504 TATGGGTCTTTGAAGATCTCAAAATCAGGGATTTGTTGAACGAAAGTGGTGTAGCGAC 563
445 TTTAGGACTCTAGAGACTCCAAATTTGAGATCTTTTGAAGAAGCTGGAGTGTAGTGC 504

564 GCGAGGTTAAGTCTGAGGTTGAGAGCTTCTGGGAAGAGCGGAAGAAAGTTGAGAG 623
505 ATCAAGAGTGAATTCAGAGGTAGAGAACTTTAGAGGAAGGAAGGAAGTTCGAAG 564

624 TGCTTCAGGGGACACAAATTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCA 683
565 TGCTTCAGGGGACACCAATTCAGCACTCAACACTTATGGCGGTGATCTTGTGGAACA 624

684 AGCAGGAAGCTTGATCTCTGATTTGCTGTGATGAGAGATTTAGAGAGTCTGTAGGAT 743
625 AGCAGGAAGCTTGATCTCTGATTTGCTGTGATGAGAGATTTAGAGAGTCTGTGAGAT 684

744 TCTTTCGAGGAGACGAGAAACAACTCTGTGCTTATTTGAGAGCCAGGAGTTGGTAAAC 803
685 TTTATCAAGGAGGACTAAGAACCAACCGGTTCTTATTTGGAGAGCCCGGTGTGGGTAAAC 744

804 AGCTGTGGTTGAAGTTTAGCACAAGGATTTGTGAAGAGATGTGCCCCAACAGTCTTAC 863
745 AGCATGTTGTTGAAGGCTAGCAGAGAGATTTGACGTGTGATGTTTCCAGATTAATTAGC 804

864 TGATGTGAGATTAATTTGTTGAGCATGGGTGCTGTAGTTGCTGGTCTTAAATPACGAGG 923
805 TGATGTTAGGCTTATGCAATTGGATATGGAGCGCTAGTTGCTGGAGCTAAAGTACAGAG 864

924 AGATTTGAAGAAAGGTTGAATCTGTTTGAAGAAGATTTGAGAGCGCTGAGGCAAGT 983
865 TGAATTTGAAGAGGCTGAAGGCTGTGCTGAAAGAGATTTGAAGAAGCGGAGGAAAGT 924

QY 984 GATTCTCTTTATTGATGAGATTCAATTTGGTCTTGGCAAAACTGAAGGTCGAT 1043
DB 925 AATATCTTTTCATTGACGATACATTTAGTCTCGTCTGGTCGGACAGAGGTCAT 984

QY 1044 GGATCAGCTAATCTGTTCAAGCCCATGTTAGCTAGAGGGCAGCTTCGATGCTTGGTC 1103
DB 985 GGATGCTGCTAATCTGTTTAAAGCCAAATGACGAGAGGTCATTTACGGTGCATTGGTGC 1044

QY 1104 TACAACGCTTGAGAAATACAGGAATATTTGAGAAAGATGCTGCTTTGAGAGAGGTT 1163
DB 1045 AACTACATCTGAGGAGTACAGGAAGTATGTTGAGAAAGATGCTGCAATTTGAGAGCGGTT 1104

QY 1164 CCAACAGTCTATGTTGGGAGCCAAAGTGTGCTGACACCAATTTAGTATCTTTAGAGGACT 1223
DB 1105 CCAGCAGGTGTATGTTGCTGAGCCTAGTGTGCTGACACTATTAGTATCTCCGTTGGT 1164

QY 1224 CAAAGAGAGTATGAGGACATCATGTTGTGCGAATCCAAACAGACAGAGCTCTTATAAATGC 1283
DB 1165 GAAGGAGAGGTATGAAGGCTCATGTTGTCAAAATTCAGGACAGAGCTCTTTGTAGTGGC 1224

QY 1284 TGCTCAGCTGTCTGCTGCTTACATTAACCTGCTGGCAATTTACGGATTAAGCAATTTGATT 1343
DB 1225 TGCCAGGCTCTCATCTCGGTACATTTACAGGTGCAATCTGCCAGATTAAGGCTATTGACCT 1284

QY 1344 GGTTCATGAGGCTTGTGCAATGTGAGAGTCCAGCTTGATAGTCAACCTGAAGAGATTGA 1403
DB 1285 AGTTGATGAAGCTTGTGCAATTTGAGAGTTCAGCTTGATAGTCACTGAGGAAATTTGA 1344

QY 1404 TAACTTTGAAGAGAGAGATGAGCTGGAATTTGAATTTCAAGCTTTGGAAGAGGAA 1463
DB 1345 CAATCTGAGAGGAGAGAAATTCAGCTAGAGGTTGAATTTCAAGCTCTCGAGAAAGGAAA 1404

QY 1464 GGATTAAGCCGACCAAGCTCGACTTATAGAGTCCGGAAGAGCTTGATGACCTGAGAGA 1523
DB 1405 AGCAAGCTAGCAAGCACGCTCTCATGAAGTGAAGAAAGATAGATGAGTGC 1464

QY 1524 CAAAGCTTCAGCTCTCACATGAAATA CAGAAAGAGAGAAAGAGAAATTTGATGAGATTGC 1583
DB 1465 CAACTCCAACTTTGATGATGAGGTACAAGAAAGAGAGAAAGGATAGATGAGTGC 1524

QY 1584 AAGGCTTAAACAGAAAGAGAGAGCTCATGTTTCTTTGAGAGGAGCAAGCAAGATA 1643
DB 1525 CAGGCTCAAGCAAAAGCGGATGAGCTCATCTATGCTTTACAAGAGCTGAAGAGGATA 1584

QY 1644 TGACCTTCAAGAGCTGCTGATCTAAGATATGGCAATTTCAAGAGTGGATCTGCAAT 1703
DB 1585 TGATCTGGCGGGGAGCAGATCTGAGATATGGGCAATTTCAAGAGTGGAACTGCAT 1644

QY 1704 TGCCCAACTTGAAGAACTTCTTCTGAGAGAAATGATGCTCAAGAAACCTTTGGGCC 1763
DB 1645 AGCAATCTTGAAGTACCTCAGCTGAA --AGTACAATGCTAACAGAGACTGTGGGTCC 1701

QY 1764 TGACACATCTGCTGAGGTGTGAGCGGTTGACAGAGGATTTCCAGTACAGACTTGGCCA 1823
DB 1702 TGATCAGATTGCGGAAGTTGTGAGTCTGTTGAGTGGTATTTCCGCTCTCAAGGCTTGTCA 1761

QY 1824 AATCAGAGAGAGGTTGATTTGTTGCTGATAGTGTGATGAGCGGTTGTGGGACA 1883
DB 1762 GAATCAGAAAGAGAACTGATTTGGTCTTGGCGATAGATTGCCAAAGAGTGGTGGGCA 1821

QY 1884 GAATCAAGCGGTAAATGACAGTTTCTGAGGCAATTTCTAAGGTCAAGGGCAGGACTTGGTAG 1943
DB 1822 AGATCATGAGTTAGAGCTGTTGCTGAGCGCTGTTAAGGTCCAGAGCTGGTTTAGGAG 1881

QY 1944 GGCACACAGCAACTGATCATTTCTTATTCCTTGGACCAACTGGTGTGGCAAACTGA 2003
DB 1882 GCCACAGCAAACTGGTTCATTTCTTTCTGGGCGCAACTGGTGTGGAAAGACAGA 1941

QY 2004 GCTCGCCAGGCTCTGCTGAGCAGCTGTTTGTGATGATGAACCTCTTAGTTCGATTTGA 2063
DB 1942 GCTCGCTAAAGCTTTGACAGAGAGCTCTTTGATGATGAATAACTGATGATCAGATAGA 2001

2064	TATGTCGGAAATATATGAGAACACACTCTGTCTCTCGCCTCATTTGGGGCAACAACGAGGTA	2123
2068	CATGTCGGAGGTACATGGAAACAACACTCTGTTTTCCGGGTGATTGGTGTCTCCACGAGTTA	2061
2124	TGTTGTTGCACGAGGAAGTGCGACAACTAACTCAGGCTGTGAGGAGGCGACCTTATTGTGT	2183
2062	TGTTGGGCAATGATAGGGAGAGACAATTACTGAGCTGTTAGGAGGCGGCTTTACAGTGT	2121
2184	CATACTCTTTTGATCAAGTGGAGAAGGCTCATGTTGCTGTCTTCAACACTCTGTCTCAAAGT	2243
2122	TGTGCTATTTTGATGAAAGTCAGAAAAGCCACTCTGCAGTGTTTAATACATATGCTTCAAAGT	2181
2244	TTTGATGATGGTTCGATTGACAGCGGCCAAGCAGGACAGTTCGNATTTTCAGGAACCTCGGT	2303
2182	TCTGATGATGGAAGGTTAAACGATGGTCAAGGCCGACAGTGAATTTACCAAACTCCGT	2241
2304	GATMAATCATGACATCAAAACCCTTCGTCTGAAACACCTCCCTGCAGGCGCTAACTGGGAAAGT	2363
2242	GATTATTATGACTTCAAACTTGGGAGCAGAGTATCTGTGTCTGGATTAAATGGGCAATG	2301
2364	AACAATGGAAGTGGCCCGGACGTGTGTGATGCGGAGGTGAGGAAAAACACTTCAGACCGA	2423
2302	TACCATGGAGACAGCTCGTGAATAATGGTCTATGCAGGAGGTGCGAAAGCAGAGTTTAAGCCCGA	2361
2424	GCTCTTGAAACAGGGCTTCACAGAGATGTGTGTGTTTCGACCCCTTTTCACATGACCACTGTGAG	2483
2362	GCTGCTGAATCGGCTGGATGAGATGTGTGTGTTTGAATCTCTGTCTCCCAAGACAGTTGAG	2421
2484	GAAAGTAGCTCGGCTTCAAATGAAAGACGTTGTGTTGTCGCGCTTGCTGAAAGAGAGTTGC	2543
2422	GCAAGTATGCGCTACCAAGATGAAGACGTTGCACCTACGGCTGGCTGAGAGGGGTATTTCG	2481
2544	TTTGGCAGTCACTGATGCTGCTTTGGACTATATCTTTGGCAGAGAGTTATGACCCGGTGTA	2603
2482	ATTGGCGCTTTACTCAGGCAGCTCCAGATGTGTCATCTCACAGAGTTTATGACCCGGTTTA	2541
2604	TGSGCTCAGGCCTATTAAGGAGATGCGATGGAGAGAAGGTGGTGACAGAACTGTCAAAGAT	2663
2542	TGSGTCAAGACCTATTATGAGATAGTGTGGAGAGAAGGTGGTGACGAGCTATCCAAGAT	2601
2664	GGTTGTGCGTGAGGAAATTCGATGAAAACTCCACTGTTTTCATAGATGAG-----GCGC	2717
2602	GCTCGTGAAGGAGGAGATTGATGAGAACTCAACGGTTTTACATAGATGCTGGGGTCAGCGG	2661
2718	TGGTGAATCTTGTGTACCGGTHAG---AAAGTGAAGTCTAGTGGACGCTTCAACAGGCAA	2774
2662	GAAAGATCTAACTACAGGGTGGAGAAGATGGAGGCGCTTGTGAATGCTGCCACCGGGCA	2721
2775	GAAAGTCAGATGTGCTGATTTCATATTGCTTAACGGGCCCAAGAGAGAAGTGTATGCAGGCTCAGGC	2834
2722	AAATCTGATATATTGATTGAGCTTCTTAATGTTGCC--CAGGAGTGATGCTGTCCAGC	2778
2835	GGTGAAGAAGATGAGGATCGAGGAAATAGAGAATGACGATATATGAGGA	2882
2779	AGTCAAGAAGATGAGGATTTGAAGAAATTTGAGATGACGAAATGGAAGA	2826

RESULT 7

66070

AAI66070 standard; DNA; 3084 BP.

AAI66070;

14-JAN-2002 (first entry)

Zea mays heat shock protein HSP101 gene (GenBank: AF133840).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton; canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato; *Arabidopsis thaliana*; ds.

Zea mays.

PN	WO200170929-A2.	
XX		
XX	27-SEP-2001.	
XX		
XX	20-MAR-2001; 2001WO-US08836.	
XX		
XX	20-MAR-2000; 2000US-190769P.	
PR	18-APR-2000; 2000US-198116P.	
XX		
XX	(ARCH-) ARCH DEV CORP.	
XX		
XX	Lindquist S, Queitsch C, Vierling E;	
XX		
XX	WPI: 2001-639123/73.	
DR	P-PSDB; AAMS1669.	
DR		
XX	Transgenic plants with improved heat stress tolerance, useful for	
PT	producing animal feed, oil and synthetic products -	
XX		
XX	Claim 4; Page -; 91pp; English.	
XX		
XX	The invention relates to a transgenic plant, comprising a genetic	
XX	construct comprising a promoter operatively linked to a nucleic acid	
CC	sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP)	
CC	family amino acid sequence (AAMS1651-AAMS1671). The transgenic plant has	
CC	increased stress tolerance, especially to heat. The plant is a cereal,	
CC	grass, ornamental plant, crop plant, food plant, oil-producing plant, a	
CC	synthetic product-producing plant, an environmental waste absorbing	
CC	plant, an alcohol plant, a medicinal plant, a recreational plant and/or	
CC	an animal feed plant. In particular, the transgenic plant is cotton,	
CC	canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or	
CC	Arabidopsis thaliana. The plants may be used to produce animal feed,	
CC	alcohol, crop, oil, medicine or a synthetic product.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained from Genbank using the Accession Number	
CC	reference provided in the specification.	
XX		
XX	Sequence 3084 BP; 670 A; 874 C; 1057 G; 483 T; 0 other;	

Query Match	43.3%	Score 1345;	DB 22;	Length 3084;
Best Local Similarity	69.1%	Pred. No. 0;		
Matches 1904; Conservative	0;	Mismatches 835;	Indels 18;	Gaps 4;

141	GGATTACAAAGCTTAATCGAAGATGAATCCAGAGAAATTACACACAGACAAACGAGAC	200
142		
143		
123	GGAGTAGCAAGAGACGACCATGAATCCGACAACTTCACCCAAGACGAACGAGGC	182
124		
125		
201	AATTGCTACAGCTCATGAGCTAGCTGTGAATGCAGACATGCTCAATTCACTCCCTTTGCA	260
202		
183	GATCTGTGGGGCGCAGAGATTGGGGTGAGGCGCGGCCACGCGCAGCTCACGCCGCTGCA	242
184		
261	TTTAGCTGGTGTCTTGTATCTCATCCACCCGATATATTCTCAAGCAAACTCTCTAGTGC	320
262		
243	CCTGCCCGCAGTGTGGCTCGCAGCAAGGGCGGCATCTCTGCGCAGGCCCATCACGGGGC	302
244		
321	CGGTGGCGAGAAACGACG-----TCAATCTGCTGAAGAGTGATCAATCAAGCGCTTGAA	374
322		
303	GTCGGGGGGCGACGGAGCGCGGGGACTCGTTGAGCGGTTGCTGAACAACTCGCTCAA	362
304		
375	GAAGCTTCTCTTCACAATCTCTCCACCTGATGATATATCCAGCGAGTTCTAGTCTTATTAA	434
376		
363	GAAGCTGCGGTGAGTCTCCGCGCGGACTCCGTCCCGGGGTCCA CGGCACATGATCAA	422
364		
435	GGTCAATCGTCTGCTCAAGCTGCTCAGAACTCAGAGGTGATATCAATTTGGCGTGTGA	494
436		
423	GGTCAATCGCGGGGGCAGTCCGCGCAGAAAACGCGGGGACTCTCACCTCGCGTCTGA	482
424		
495	CCAGTTGATTAATGGGTCTTTTGAAGATTCTAAATCAGGGATTTTGTGTGAACGAAGTCGG	554
496		
483	CCAGCTGTGCTCGGCTGTCTGAGGACTCGCAGATCTCCGATGCTCCAGTGGCGCGG	542
484		
555	TGTACGCGACGGCGGGTAAAGTCTGAGTTGTAAGAAGTTCGTGGGAAGAGAGGGAAGAA	614
556		

543	CGTGTCCGCGCGGGTGC	CGCGCGAGCTTTG	GAAAGCTCCGCGCGCGGGAGGCGCGCG	602	
615	AGTTGAGAGTGTCTT	CAGGGGACACAAATTTT	CAAGCTTTAAAGACTTATG	GAAGAGATTT	674
603	CGTGAGTCCGGT	CGGGGACACCACTT	CCAGGCGCTCA	GACATACCGCGGACCT	662
675	GGTTGAGCAAGCAGGGAAGCTGTATCT	GTGATTTGGTCTGTGATGAGGAGATTTAG	AAGAT		734
663	CGTTGAGCAGCGCGGAAAGCTGGACCCCGT	CATCGCGCGGACGAGGAGATCCGCGCGT			722
735	CGTGAGGATCTTTTCGAGAGAACGAGAA	CAATCTCTGTCTTATTTGAGAGCCAGGAGT			794
723	GGTGGCGATCTCTCGCGCGGCACTTA	GAAACAACCCCGTCTCATCGCGAGCCCGCGT			782
795	TGCTAAAAACAGCTGTGGTTGAAGGTTT	TAGCAACAAGGATTTGAAAGAGATGTGCCCAA			854
783	TGGCAAGACGGCGTGTGGAGGGCTCGCG	CAGCGCATCGTTCCGCGCAGCTGCCCCAG			842
855	CAGCTTACTGATGTGAGATTAATTTCT	GTGGACATGGGTGCGTTAGTTGCTGTGTCTAA			914
843	TAACCTCTCGAGCTCCGCTCATCGCGCT	CGACATGGCGCGCTCGTCGCGCGGCCAA			902
915	ATACCGAGGAGGTTTGAGAAAGGTTGAAATCT	GTGTTTGAAGAACTTGAGGACGCTGA			974
903	GTACC	CGCGCGAGTTTCGAGGACGGGCTCAAGGCGGTGCTCAAGGAGGTGGAAGAGGCCGA			962
975	AGGCAAGTGATCTCTTTATTTGATGAGATTCATTT	TGGTCTCGTGGCAAACTGA			1034
963	GGGGAAGTCACTCTTCTNTCGACGAGATAC	CACTCGTCTGGCGGGCGGAGGCGGA			1022
1035	AGGTTCGATGGATGACGTAAATCTGT	TTCAAGCCCATGTTAGCTAGAGGCGAGCTTCGATG			1094
1023	GGGTTCCATGACCGCGGCCAACTGTT	CAAGCCCAATGTTGGCGAGGGGACAGCTTCAGGTTG			1082
1095	CATTGGTCTCAACCGCTTGAGAAATGAGAAAT	TGTTGAGAAAGATCTGCTTTGA			1154
1083	CATCGCGCGCACCACTGGTGGAGGAGTACCG	CAAGTACGTGTGAGAGAGACGACGGTTTGA			1142
1155	GAGGAGGTTCCAAAGTCTATGTTGCGGAGCCAA	GTGTGCTGCACACCAATTAGTATCTCT			1214
1143	CGCGCGGTTCCAGAGGTGTTGCTCGCGAGCGGAGCGT	GCCGACACCGTCAGCATCT			1202
1215	TAGAGACTCAAGAGAAAGTATGAGGGA	CATCATGTGTGCGAATTC	CAAGACAGAGCTCT		1274
1203	GAGGGACTCAAGGAGAAAGTACGAGGGG	CAACATGCGGTGAGGATCCAGGACCGCGCCT			1262
1275	TATAAATGCTGCTCAGCTGTCTGTCTGCTT	CAATTAACCTGGTGGCATTTACCGGATTAAGC			1334
1263	CGTGTCCGCGCACAGCTATCCGCGAGGTAT	CCGCGAGGTACATCATGGGTGCGCACCTGCTGTGCAAAAGC			1322
1335	AAATGATTTGGTTGATGAGGCTTGTGCGAA	TGTGAGAGTCCAGCTTCATAGTCAACCTGA			1394
1323	CATAGACTGGTGGACGAGGCTGCGCAATGT	GNGGTTGAGCTTCGACAGCCAGCCGGA			1382
1395	AGAGATTGATAACCTTTGAAAGGAGAGAGAT	GCAGCTGGAAATGAAATTTCA	CGCTTTGGA		1454
1383	GGAGATTGATAACCTCGAGAGAGAA	GAGAAATCCAGCTTGAGGTTGAGCTCCAGCGCTCGA			1442
1455	AGGGAGAGAGATAAGCCAGCAAGCTG	CACTTATAGAGGTGCGGAAGAGCTTGATGA			1514
1443	GAAGGAGAGGACAAAGGCGAGCAAAAGCCCGGCT	GTATCGAGGTTCAGGAGGAATTTGGACGA			1502
1515	CTTGAGAGACAAGCTTCAGCTCTCA	CGATGAAATACAGAAAGGAGAAAGAGAAATTGA			1574
1503	TTTGAGGGACAAGCTGAGCCCTGACCAT	GAATGAGTACAGGAAGAGAGAGAGATTCGA			1562
1575	TGAGATTGGAAGGCTTAAACAGAAAGAGAA	GAGCTCATGTTTTTTCTTTGCAAGGAGGCGGA			1634
1563	TGAGATTCAGGAAGCTGAGCAGCGCGGAGGAG	CTCCAGTTTCACTCTGCAAGGAGGCGGA			1622
1635	ACGAGATATGACTTTGCAAGAGCTGTGAT	TCTTAAGATATCGGCGCAATTCAGAAAGTGA			1694
1623	CGCGCGGATTGACCTGGCCGCTGTGCGCGAG	CACTCAAGTACGCGCGCTTCCAGGAAATTCGA			1682
1695	ATCTGCAATTTCCCAACTTTGAAGGA	ACTTCTTTGAAAGAGATTTGTAAGAGATTTGTAATG	TCTCA	CAGAAAA	1754
1683	CGCGCGGATCTCCAAGCTGGA	---GAGCAAAACAGGGGAGAACTGATG	TCTCACCGAAAC		1739
1755	CGTTGGGCTGAAACATCTGCTGAGGTTGT	GAGCGGTTGGACAGGGATTTCCAGTGACGAG			1814
1740	CGTCGGCCCTGAAACAAATTCAGAGGT	TGGTGAAGCGGTTTGAAGCGGATTTCCAGTGACCGG			1799
1815	ACTTGGCCAAAATCAGAAAGAGAGAGT	TGATTTGGTCTTGCTGATAGGTTGCATTAAGCGGGT			1874
1800	GTTTGGCCAGAAACGACAGAGAGAGG	CTGGCTGGCTGACAGGCTTCAACNAGAGGGT			1859
1875	TGTGGGA	CAGAAATCAAGCGGTAAATG	CAGATTTCTGAGGCAATTTCTAAGGTCA	AGGGCAGG	1934
1860	GGTCGGCCAGACAGAGGCTGTGAGCGCG	CTGCGAGAGCGGTGCTGAGGTCGAGGGCGCG			1919
1935	ACTTGGTAGGSCACAA	CAGCCAACTGGAATCATTTCTTATTTCTTGGACCAACTGGTCTTGG			1994
1920	TCTTGGCAGGCCAACACAGCCCACT	GGCTGGTCTCTTCTGGTCCGACTGGCGTGGG			1979
1995	CAAAACTGAGCTCGCCAAAGGCTCT	TGTGAGCAGCTGTTTGATGATCAAAACCTCTTAGT			2054
1980	GAAAACTGAGCTAGCCAAAGGCCCT	AGCGAAACAGCTGTTGACGACGAGAACTTCTTGT			2039
2055	TCGGATTGATGTGCGGAATATATGGA	ACAACTCTGTCTCTGCTCTTGGGCAACC			2114
2040	CGGCATCGACATGTGCGAGTACATGG	AGCAGCACTCGCTTCTCGCTCATCGGAGCACC			2099
2115	ACCAGGTATGTTGGTCACGAGGAA	GTTGGACAACTAACTGAGGCTGTGAGGAGGCGACC			2174
2100	GCCTGGCTACGTCGCGCCATGAAG	AGGGTGGGCACTGACTGAAACAAGTGAAGGAGGAGGCC			2159
2175	TTATTGTGCTACTCTTTGATGAAGT	TGGAAGGCTCATGTTGCTCTTCAACACTCT			2234
2160	GTACAGCGTGATCTGTTTCGACGAG	GTGAGAGGCCCATGTGCGCGTGTCAACACCCT			2219
2235	GCTCCAAAGTTTGGATGATGTGCA	ATTGACAGACGGGCAAGGACAGTGCATTTTCAG			2294
2220	GCTCCAGGTCTCGACGACGCGAG	TGTGA	CGGATGGGCAAGGACGCGTGGACTTCAG		2279
2295	GAACTCGGTGATATCAT				

223 GCTGTGAATGCAGGACATGCTCAATTCACTCTTTCGATTAGCTGGTGCTTTGATCTCT 282

1234	TCGAGGTACATCATGTTGGTGGCCCATTTGGCTCTGATTAAGCAATCGAATTTGGTTGACGAGGCC	1293	CTTGACGAGATTGTTGGTGTTCGACCCCTTTTCACATGACACGAGTTGAGGAAAGTACGCTCGG	2499
1357	TGTGCGAATGTGAGAGTCCAGCTTGATAGTCAACTGTAAGAGATTGATTAACCTTTGAAAGG	1416	CTGAGCAGAGATCGTTCATCTTCGACCCCTCTGTTCGATGAGCAGCTGCGGAAAGGTCGCTCGG	2427
1294	TGCGCAATGTGAGGCTGCACTTGACAGCAGCCTGAAGAGATCGACAACCTCGAGAGG	1353	CTTCAAATGAAAGACGTTGCTGCTCCGGCTTCGCTGAAGAGAGAGTTGCTTTTGGCAGTCACT	2556
1417	RAGAGATCGAGCTGGAATTTGAACTTACGCTTTGGAAGGGGAGAGATTAAGCCAGC	1476	CTTCAGATGAAAGATGTGGCAGTCCGCTTTGCCGAGAGGGCGTTGCTCTGGCCGTCACC	2487
1354	AAGAGATCCAGTTGGAATGAGCTCCATGCCCCCTCGAAGAGGAGAGCAAGGCTAGC	1413	GATGATCGACGAACTCCACGCTGTACATCGACGCTGCGCCACGAGGACGAGCTGACC	2667
1477	AAAGCTCGACTTATAGAGTGGGAAAGAGCTTTGATGACCTCGAGAGACAAGCTTCAGCCT	1536	GAAATCGATGAAACTCCACTGTTTACATAGATGACG-----CGCTGGTGAATCTTTGTG	2730
1414	AAAGCTCGCTAGTTGATGTGAGGAAGAGTTGGACGATCTGAGAGACAAGCTCGAGCG	1473	ATCCGAGATGATCGAGAGAGGATGATGACGAGCTCTCCAGATGTTGATCCGCGAG	2607
1537	CTCAGATCAATACAGAAGAGAGAGAGAGATTTGATGATTCGAGAGCTTTAAACAG	1596	ATAAGAGATGATCGAGAGAGAAAGTGTGTGACAGAACTGTCAAGAGATGTTGTGCGTGAG	2676
1474	CTCAGATGAAGTATCGCAAGAGAGAGAGAGGATTCGACGAGATCAGGAGCTGAAGCAG	1533	GAAATCGATGAAACTCCACTGTTTACATAGATGACG-----CGCTGGTGAATCTTTGTG	2730
1597	AAAGAGAGAGCTCATGTTTCTTTGACGAGGAGAGAGAGATATGACTTTGCAAGA	1656	GATGATCGACGAACTCCACGCTGTACATCGACGCTGCGCCACGAGGACGAGCTGACC	2667
1534	CGCGCGAGAGCTGCAGTTCACTCTGACGAGGCGGAGCGCGGATGGAATTTGGCCCGC	1593	TACCGGTTAGAAA---GTGGAGGTTCTAGTGAACGCTTCAACAGGCAAGAGTTCAGATGTG	2787
1657	GCTGCTGATCTAAGATATGCGGCAATTTCAAGAGTGGAACTCTGCAATTTGCCACTTGA	1716	TATGGCGTCGACAAGCAAGGAGGCTGTGTGAACGCGCACGCGGCCCAAGTCCGACATC	2727
1594	GTGGCTGATCTCAGATACGCTGCTCTGACGAGAGTTGACGCTGCTCAATTTGCTAAGCTG	1653	CTGATTCATATTGCTTAAACGCGCCAAAGAGAGTGTATGATGATGATGATGATGATGATG	2847
1717	GGAACTTCTTCTGAAGAGAACTGTGATGCTCACAGAAAAGTTGGGCTCTGAACATTTGCT	1776	CTGATCCAGGTTCTTAGCGAGCTGTTGGGGCGGATGCGGCGCACGCGCTGAAGAAGATG	2787
1654	GG---TGAGACCGCGGAGAAATTTGATTTAAACAGAGACTGTGCGGCCCGACGATGCT	1710	AGGATCGAGGA 2858	
1777	GAGTTGTGAGCGGTTGGACAGGATTTCCAGTGACGAGACTTTGGCCAAAATGAGAAGGAG	1836	AGATCATGCA 2798	
1711	GAGTGTGAGCGCTGGACTGCTATTTCTGTCACAGGCTCGGACAGAAATGAGAAGGCG	1770		
1837	AGTTGATGTTGCTCTGATAGTGTGCATACGCGGTTGTGGACAGAAATCAAGCGGTA	1896		
1771	AGGCTGATCGGGCTGCGAGATCGACTGCAATCAGAGGGTGTGTGGACAGTATGAGCGGTC	1830		
1897	AATGCAAGTTCTGAGGCAATTTCTAAGTCAAGGCGAGGACTTTGGTAGGGCAACACAGCCA	1956		
1831	AATGCAAGTTGAGAGAGCTGTTCTACGTTGAGGCTGTGCTTTGGACGCGCCGACGAGCT	1890		
1957	ACTGATCAATCTTATTCCTTGACCAAACTGTGTGTGGCAAACTGAGCTCGCCAGGCT	2016		
1891	ACTGTTTCACTTCTTCTTCTTGACCGACGCGGTGTGGAAAACCGAGCTCGCCAAAGCT	1950		
2017	CTTGCTGAGCAGCTTTGATGATGAACCTTTAGTTGCGATTTGATATGTCGGAATAT	2076		
1951	CTAGCTGAGCAGCTTTGATGACGAGAACCTTGCTGCTCGCATCGACATGTCTGAATAC	2010		
2077	ATGGAACAACACTCTGCTCTCGCTCATTTGGGGCACACACAGGATGTTGGTCAAGAG	2136		
2011	ATGACAGCAATTCGGTTGCGCGCTTAATCGGAGCGCCACCTTGGATATGTTGGTCAAGAA	2070		
2137	GRAGGTGGAACAATCACTGAGGCTGTGAGGAGGCGACCTTATTTGTGTCATCTCTTTGAT	2196		
2071	GAAGGGGGGAGCTGACCGAGCAAGTGAAGGAGGAGGCCATACAGTGTCTATCTCTTCGAC	2130		
2197	GAAGTGGAGAAGCTCATGTTGCTGCTTCAACACTCTGCTCCAAAGTTTGGATGATGCT	2256		
2131	GAGTTGAGAGGCGCATTTGGCGGTGTTCAACACTCTGCTCCAGTCTCGACGATGGG	2190		
2257	CGATTGACAGCGGCAAGGACGAGCAGTGTGATTTGAGAACTCGGTTGATATCATGACA	2316		
2191	CGGTTGACCGAGGCAAGGACGAGCGGTTGATTTTGGAGAACACCGGTGATCATCATGACC	2250		
2317	TCAAACTTGTGCTGAAACACTTCTTGGAGGCTTAACCTGGGAAAGTAACTGAAGTGTG	2376		
2251	TCAAACTTGGCGGAGGACCTTCTCGCGGATGTTGGGCAA---TTGATGAGGTT	2307		
2377	GCCGGGACTGTGTGATGCGGGAGGTTGAGGAAACAATTCAGACACGAGGCTCTTGAACAGG	2436		
2308	GCTCGTATCTGTTCTATGACGAGGTTGAGGAGCAATTCGCGCCGAGCTGTGTAACCGT	2367		

sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAI66057-AAI66084). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 2821 BP; 524 A; 902 C; 1035 G; 360 T; 0 other;

Query Match 39.8%; Score 1236.2; DB 22; Length 2821;

est Local Similarity 67.3%; Pred. No. 0;

atches 1847; Conservative 0; Mismatches 868; Indels 30; Gaps 6;

156 ATCGAGATGATCCAGAGAAATTCACACACAGACAAACAGAGACAAATGCTCAGACTCA 215

58 AGCGCCATGAACCCGGAACATTCACGCAACAGACCAACAGAGCGCTGCTGGCGGCGCA 117

216 TGAGCTAGCTGTGAATGACGACATGCTCAATCTCACTCTTTTGCATTTAGCTGTGCTTT 275

118 CGAGCGCGCTCGAGGCGCGCCACGCGCAGATCACGCGTTTGCACCTGGCGGCGGCT 177

276 GATCTGTATCCACCGGTATATTTCTCAGCAATCTCTAGTCCGCGGTGGCGGAGAA--- 332

178 GCGCGCGGACAAAGTCGGGCGCTCTCGCGCGCGCGCTCGCGGCGGCGCTCCGCGGGAATGC 237

333 ---CGCAGCTCAATCTGCTGAAGAGTGAATCAATCAAGCTTTGAAGAGCTTCTCTTCA 389

238 GTCCCGGGGAGCTGCTCGAGCGCGTCTCGCGCGCGCTCAGGAAGCTGCGCTCGCA 297

390 ATCTCTCCACCTGATGATATTCAGCGAGTTCTAGTCTTATTAAGGTCTATTCGTGTCG 449

298 GTCCCGCGCGGACTCGTCTCGCGCTCCACGCGCTCATCAAGGCTATCCGCGCGCGC 357

450 TCAAGCTCTCAGAGTACGAGGTGATCTATTTGGCTGTTGACCAAGTTCATTTATGGG 509

358 GCAGTCGGCGGAGAGGAGCGCGGAGCTCGCACCTCGCGCTCGCACCAAGTCTCATGGG 417

510 TCTTCTGAAGATTCCTCAATCAGGATTTGTTGAACGAGTTCGCTGAGGACGCGGAG 569

418 CTCTCTCGAGGACGCGGATCGCGGCTCGCTCAAGGAGCGCGGCTCGCGCTCGG 477

570 GGTAAAGTCTGAGGTTGAGAGCTTCTG---GGGAAAGAGGGAAGAGTTGAGAGTGC 626

478 GGTGCGCGCGGAGCTCGAAGCTCGCGCGCGGAGCAACTCGCGCAAGTCTGAGTCCGC 537

627 TTCAGGGGACACAAATTTCAAGCTTTAAGACTTATGAAGAGATTTGGTTGACCAAGC 686

538 CTTCGCGGACACCACTTCAGCGGCTCAAGAGTACGCGCCGCGCTCTGTCGAGGTGCG 597

687 AGGGAAGCTTGTATCTCTGTTGATGAGGAGATTTAGAGAGTCTGAGAGATTTCT 746

598 CGGCAAGCTCGACCCCGTCTATCGCGCGGACGAGGAGATCCGCGCGCTGTCGCTCCT 657

747 TTCGAGAGAACCAAGAACATCTCTGCTTATTTGAGAGCAGGAGTTGTTGAAACAGC 806

658 CTGCGCGCGGACCAAGAACCAACCCCGCTCTATTCGCGGAGCGCGGCTCGGCAAGCGC 717

807 TGTGGTTGAGGTTTATGACCAAGAGATTTGAAAGAGAGATGTCGCCAAGCTTTACTGA 866

718 CGTCTCGAGGCGCTCGCGAGCGCGCTGTCGCGCGGAGCTGTCGCGCAACCTGCTCGA 777

867 TGTGAGATTAATTTCTGTTGACATGCGGTGCTTGTGCTGCTGCTTAAATACCGAGGAGA 926

778 CGTCCGCTCTGCTCGCTCGACATGAGCGCGCTCTGCTGCGCGCGGCGCAAGTACCGGCGGA 837

927 GTTTGAAGAAAGGTTGAAATCTCTTTTGAAGAGAGTTGAGGAGCGCTGAAGGCAAGTGTAT 986

838 GTTTCGAGGAGCGCTCAAGCGCGTCTCAAGGAGGTGGAGGAGCGCGCAAGGTCAT 897

987 TCTCTTTATGATGATGATTTCTTCTGCTCTGCAAACTGAAGGCTCATGCA 1046

898 CTTCTTCATCGACGAGATACACTCTGCTCTGCGCGCGGAGGACGAGGGGTTCATGGA 957

1047 TGCAGCTAATCTGTTCAAGCCCATGCTAGCTAGAGGCGAGCTTCGATGCTGCTGCTAC 1106

958 CGCGGCCAACCTCTTCAAGCCCATGCTGCGCAGGCGCGAGCTCCGGTGCATCGCGCCAC 1017

1107 AAGCTTTGAAGATACAGGAATATGTTGAGAAAGATGCTGCTTGGAGGAGGCTTCA 1166

1018 CACGCTGGAGGAGTACCGCAAGTACGCTGAGAAAGACGCGCGTTCGAGCGGACGCTTCA 1077

1167 ACAAGTCTATGTTGCGGAGCCCAAGTGTGCTGACACCATTTAGTATCTTTCAGAGGACTCAA 1226

1078 GCAAGTGTGCTGCTGCGGAGCGCGCTGCGCGACCATCAGCATCTCTCGAGGCTCAA 1137

1227 GGAAGATGATGAGGACATCATGCTGCTGCGCATTTACCGGATTAAGCAATTTGATTTGCT 1286

1138 GGAAGATGATGAGGAGCACACGCGCTGCGCATCCAGGACCGCGCATCTGCTGCTGCGCGC 1197

1287 TCAAGTCTGCTGCTGCTTACATTAACCTGCTGCGCATTTACCGGATTAAGCAATTTGATTTGCT 1346

1198 GCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257

1347 TCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1406

1258 GGAAGATGATGAGGAGCACACGCGCTGCGCATTTACCGGATTAAGCAATTTGATTTGCT 1317

1407 CTTGAAAGAGAGAGATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1466

1318 TCTGAGCGGAGCGGATTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1377

1467 TAAAGCGCAAGAGCTGCTGCTTATAGAGTTCGCGGAGAGCTTTGATGCTTACGAGACAA 1526

1378 CAAGGCGCAGCAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1437

1527 GCTTCAAGCTTCTCAAGTGAATPACAGAAAGAGAGAGAGATTTGATGCTGCTGCTGCTGCTGCT 1586

1438 GCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1497

1587 GCTTAAACAGAAAGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1646

1498 GCTGAGCAGCGCGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1557

1647 CTTGCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1706

1558 CTTGCGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1617

1707 CCAACTTTGAAGAGAGCTTCTTCTGAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1766

1618 GAAGCTGAGAGG---CGAGACCGCGGAGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1674

1767 ACATGCT 1826

1675 GCAGATCGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1734

1827 TGAGAGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1886

1735 GCACAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1794

1887 TCAAGCGGTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1946

1795 GGAGCGCTGCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1854

1947 ACAACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2006

1855 GCAGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1914

2007 CGCCCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2066

1915 CGCCAGGCGCTCGCGAGCAGCTGTTGACGACGAGAACTGCTCGTCGCATCGACAT 1974
2067 GTCCGAATATATGAAACAACTCTGCTCTCCCTCATTTGGGCGACCAACAGGATATG 2126
1975 GTCCGATATGAGCGACACTCCGTCGCTCGCTCATCGCGCGCCACAGGCTATGT 2034
2127 TGSTACAGAGAGGTGGACAACTAATGAGGCTGTGAGGAGCGACCTTATTGTGCTAT 2186
2035 CGGACACGAGAGGCGGCGAGCTGACGAGAGAGGTGAGGAGGAGCGGTACAGCGTGAT 2094
2187 ACTCTTTGATGAAGTGAAGAGCTCATGTTGCTGCTTCAACACTCTGCTCAAGTTT 2246
2095 CCTGTTTCGACGAGTGGAGAGCGCAGTGGCGGTGTTCAACACGCTCTCCAGTGCT 2154
2247 GGATGATGCTGATGACACAGCGGCAAGGACGACAGTGCATTTTCAGGAACCTCGTGAT 2306
2155 GGAGCGCGCGCTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2214
2307 AATCATGACATCAAACTTGTGCTGTAACACTCTCTTGCAGGCTTAACCTGGGAAAGTAAC 2366
2215 CATCATGACCTCAACCTCGCGCGGAGCACTCTCTCGCGGATGCTGGGCAAGAACTC 2274
2367 AATGAGAGTGGCGGCGGAGCTGTGATGCGGAGGTGAGGAAACACTTCAGACAGCT 2426
2275 CATGAAAGTTCGCGCGGAGCTCTGTCATGCGAGAGGTGAGGAGCACTTCAGGCGGAGCT 2334
2427 CTTGAAAGAGCTTACAGAGATGTTGTTGCGACCCCTTTTCATGACAGCCAGTTGAGGAA 2486
2335 GCTGAGCGGCTGACAGAGATGTTGATCTTGCACCGCTGTCACAGAGAGTGGGAA 2394
2487 AGTAGCTGCGCTTCAATGAAGAGCTGTGCTGCGGCTTGTGGAAGAGAGTTCCTTT 2546
2395 GGTGCGCGCGCTGACAGATGAAGAGAGCTGCGCGCTCGCGAGCGCGCTGCACT 2454
2547 GCAGAGTCACTGATGCTGCTTGGACTATATCTTGGCAGAGATTAAGCCGCTGTATGG 2606
2455 GCGCGCTCACGAGCGCGCTGAGAGCTGTCCTGCTGCTGCGCTGATCCAGTGTATGG 2514
2607 TGCTAGGCTTAAAGAGATGAGTGAAGAGAGGTGCTGACAGAACTGTCAAGATGCT 2666
2515 CGCTCGCGCATCAGAGGTGATCGAGAGAGGTGTTGACGAGCTGTTCAAGATGCT 2574
2667 TGTGCTGAGAAATCATGAATCAATCTCACTGTTTACATAGATGCGAGGCG---CTGGTGA 2723
2575 GATCCAGAGAGATGACAGAGATGCTGCTGATCATGACGCGCGCGGACAGAGACGA 2634
2724 TCTGTGTACCGGTAGA---AAGTGAAGTCTAGTGAGCGCTTACACAGGCAAGATC 2780
2635 GCTGCGCTACCGGTGACCGCGCGCGGCTGTTGAACGCGGAGACGCGGCGGAGAGGTC 2694
2781 AGATGTGCTGATTCATATTGCTAACCGGCGC-----AAAGAGAGTGTGATGCGC 2828
2695 CGACATCTCATCCAGGTGCGGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGG 2754
2829 TCAGGCGGTGAGAGATGAGGATCGAGGAAATAGAGATGACGA 2873
2755 CAAGCGCGTGAAGAGATGAGGCTCATGAGAGAGCGGCGGAGGCA 2799

JUL 10

16071

AAI66071 standard; DNA; 2045 BP.

AAI66071;

14-JAN-2002 (first entry)

Zea mays 101 kDa heat shock protein gene (GenBank: AF083327).

Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;
canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;
Arabidopsis thaliana; ds.

Zea mays.

WO200170929-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI: 2001-639123/73.

P-PSDB; AAM51670.

Transgenic plants with improved heat stress tolerance, useful for
producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic
construct comprising a promoter operatively linked to a nucleic acid
sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)
family amino acid sequence (AAM51651-AAM51671). The transgenic plant has
increased stress tolerance, especially to heat. The plant is a cereal,
grass, ornamental plant, crop plant, food plant, oil-producing plant, a
synthetic product-producing plant, an environmental waste absorbing
plant, an alcohol plant, a medicinal plant, a recreational plant and/or
an animal feed plant. In particular, the transgenic plant is cotton,
canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or
Arabidopsis thaliana. The plants may be used to produce animal feed,
alcohol, crop, oil, medicine or a synthetic product.
Note: the sequence data for this patent did not form part of the printed
specification, but was obtained from GenBank using the Accession Number
reference provided in the specification.

Sequence 2045 BP; 470 A; 532 C; 673 G; 370 T; 0 other;

Query Match 30.0%; Score 931.4; DB 22; Length 2045;
Best Local Similarity 72.1%; Pred. No. 2e-246;
Matches 1259; Conservative 0; Mismatches 476; Indels 12; Gaps 3;
QY 1145 CTGCTTTGAGAGGAGGTTCCAAAGTCTATGTTGCGAGCCCAAGTGTGCTGCACCA 1204
DB 1 CAGCGTTTCAGCGCGGTTCCAGAGGTGTTGTCGCGAGCCGAGCGTCCCATACCG 60

QY 1205 TTAGTATCTTTAGAGGACTCAAGGAGAGATGAGGGA CATCATGTTGTGCGAATCCAA 1264
DB 61 TCAGCATTTCTGAGGAGACTCAAGGAGAGATGAGGAGGACCATGGCGTGAGGATCCAG 120

QY 1265 ACAGAGCTTTATAAATCTGCTCAGCTGTCTGCTGTTACATACTGTCGGCATTTAC 1324
DB 121 ACCGCGCCCTCGTGGTTCGCGCACAGCTCTCCGAGGATACATCATGCGTCCGCACTTC 180

QY 1325 CGGATAAGCAATTCATTTGTTGATGAGGCTTGTGCGAATGTGAGAGTCCAGCTTGATA 1384
DB 181 CTGACAAAGCCATAGACCTGTTGACGAGGCTTGGCCCAATGTGAGGTTGAGCTCGACA 240

QY 1385 GTCAACCTTGAAGAGATTGATAACCTTGAAGAGAGAGATGAGGATCGAGCTGGAATTTCA 1444
DB 241 GCCAGCCGAGGAGATTGATAACCTTGAAGAGAGAGAGATCCAGCTTGAGGTCGAGCTCC 300

QY 1445 AGCGCTTGAAGAGAGAGATGAAGCCCAAGCTCCGACTTATAGAGTGGGGAAG 1504
DB 301 ACGCGCTCGAAGAGAGAGAGCAAGCGCCAGTAAAGCCCGCTGATTGAGTTCAGGAGG 360

QY 1505 AGCTTTGATGACTGAGAGACAGCTTTCAGCTCTCACGATGAAATACAGAAAGAGAGAA 1564
DB 361 AATTGAGCGATCTGAGGAGCAAGCTTGACGCCCTTGACCATGATGATACAGGAGGAGG 420

Db	1498	TGACGCAGCTGTGGAAGATGCTGATCCAGGAGGAGATCGACGAGAACTGCACGGTCTACA	1557
Qy	2705	TAGATGCAGCGCGTGGT-----GATCTTGTGTACCGGTAGA---AAGTGGAGGTCTAG	2755
Db	1558	TCGACGCGCGCGCGGTAGGACGAACCTGGTCTACAGGCTGCACCGGAGCGCGTCTGG	1617
Qy	2756	TGGACGCTTCAACAGGCAAGACTCAGATGCTGATTCATATGCTTAACGGGCCAAGA	2815
Db	1618	TGAACGCTGAGACGGGGATGAAGTCGGAATCCTGATCAAGTCCCAACAGCTCCACCA	1677
Qy	2816	GAACTGATGCAGCTTCAGGCGGTCAAGAAAGATCAGGATCGAGGAATAGAGATGACGATA	2875
Db	1678	GGAGCGAGCTGCGCAGCGCGTCAAGAAATCAGGATCATGGAGGAGGACGAGGACGCA	1737
Qy	2876	ATGAGGA 2882	
Db	1738	TGGACGA 1744	
RESULT 11			
AAI66072			
ID	AAI66072 standard; DNA; 2045 BP.		
XX	AAI66072;		
XX	14-JAN-2002 (first entry)		
DE	Zea mays heat shock protein 101 gene (GenBank: AF077337).		
XX	Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;		
KW	canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;		
KW	Arabidopsis thaliana; ds.		
OS	Zea mays.		
XX	WO200170929-A2.		
XX	27-SEP-2001.		
XX	20-MAR-2001; 2001WO-US08836.		
XX	20-MAR-2000; 2000US-190769P.		
PR	18-APR-2000; 2000US-198116P.		
XX	(ARCH-) ARCH DEV CORP.		
PA	Lindquist S, Queitsch C, Vierling E;		
PI	WPI; 2001-639123/73.		
DR	P-PSDB; AAM51663.		
XX	Transgenic plants with improved heat stress tolerance, useful for		
PT	producing animal feed, oil and synthetic products -		
XX	Claim 4; Page -; 91pp; English.		
PS	The invention relates to a transgenic plant, comprising a genetic		
XX	construct comprising a promoter operatively linked to a nucleic acid		
CC	sequence (AAI66057-AAI66084) encoding a plant Heat Shock Protein (HSP)		
CC	family amino acid sequence (AAM51651-AAM51671). The transgenic plant has		
CC	increased stress tolerance, especially to heat. The plant is a cereal,		
CC	grass, ornamental plant, crop plant, food plant, oil-producing plant, a		
CC	synthetic product-producing plant, an environmental waste absorbing		
CC	plant, an alcohol plant, a medicinal plant, a recreational plant and/or		
CC	an animal feed plant. In particular, the transgenic plant is cotton,		
CC	canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or		
CC	Arabidopsis thaliana. The plants may be used to produce animal feed,		
CC	alcohol, crop, oil, medicine or a synthetic product.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained from GenBank using the Accession Number		
XX	reference provided in the specification.		

20-MAR-2001; 2001WO-US08836.

20-MAR-2000; 2000US-190769P.

18-APR-2000; 2000US-198116P.

(ARCH-) ARCH DEV CORP.

Lindquist S, Queitsch C, Vierling E;

WPI; 2001-639123/73.

P-PSDB; AAM51671.

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AA166057-AA166084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AAM51651-AAM51671). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from Genbank using the Accession Number reference provided in the specification.

Sequence 3942 BP; 834 A; 1111 C; 1236 G; 760 T; 1 other;

Query Match 26.0%; Score 807; DB 22; Length 3942;

Best Local Similarity 58.5%; Pred. No. 6.6e-212;

Matches 1918; Conservative 0; Mismatches 845; Indels 514; Gaps 9;

117 CGAGTCTCATATATCGTTTAAAGGATTAACAAAGCTAATCGAAGATGAATCCAGAA 176

390 CCGCTTGTGTAGTAGTTGAGGACTGAGGACCGAGAGCAGCCATGAATCCGACAA 449

177 ATTACACACAAACAGACAAATGCTACAGCTCATGAGCTAGCTGTGAATCCAGG 236

450 CTTACACCAACAGACGACGAGGAGTCTGGGGGCGCAGATTCGGTGGAGCCGG 509

237 ACATGCTCAATCACTCTTTGCAATTTAGTGTGTGCTTTGATCTGTATCCCAACCGGTAT 296

510 CCAGCGCAGCTCAAGCGGCTGCACCTGGCGCAGTGTGCTGGCGACAAAGGGCGGCAT 569

297 ATTTCTCAAGCATCTTAGTCCGCTGGCGGAGACGACGAGC-----TCAATCTCTGA 350

570 CTTGCGGAGCCCATCAAGCGGCGTCCGGGGGCGAGCGAGCGCGGGGACTCTGTCCA 629

351 AAGAGTGATCAATCAAGCCCTTGAAGAGCTTCTCTTCAAACTCTCTCCACTGATGATAT 410

630 GCGGTCTGACAACTCGCTCAAGAGCTGCGCTGCGAGTCCCGCGCGGACTCCGT 689

411 TCCAGCAGATTTAGTCTTATTAAGTCAATCTGCTGCTGCTCAAGCTGCTCAGAGTCAAG 470

690 TCCGCGCTACGCGGCTGATCAAGTCAATCCGCGGCGCAGTCCGCGCAGAGAAACG 749

471 AGGTGATCTATTTGCTGTGACAGTGTGATTTATGSGTCTTCTTGAAGATTTCTCAAT 530

750 CCGGACTCGCACTCGCGGCTGACGACGAGTCTGCTCGGCCCTGCTCGAGACTCGGAT 809

531 CAGGATTTGTTGAACAGTCCGCTGTAGCGACGCGCGAGGCTAAAGTCTCGAGGTTGAGAA 590

810 CTCGACTGCTCAAGAGGCGCGGCTGCTCGCGCGGCTGCGCGCGGCTTGGAA 869

591 GCTTCTGGGAAAGAGGAAAGTTGAGAGTGTCTTCAAGGAGACAAATTTTCAAGC 650

870 GCTCCGCGCGGAGGCGCGCGCTGGAGTCCGCGTCCGGGATACCAACTTCCAGGC 929

651 TTTAAAGACTTATGGAAGAGATTTGTTGAGCAAGCAGGAGGAGCTTGAATCTCTGATGG 710

930 GCTCAAGACATACGCGCGGAGCTCTGTTGAGAGGCGCGGAGGCTTGAATCTCTGATGG 989

711 TCGTGATGAGGAGATTAAGAGAGTCTGTTGAGGATTTCTTTCGAGAGAACGAAACAATCC 770

990 CCGGACGAGGAGATCCGCGCGCTGTTGCGCATTTCTCTCGCGCGCACCAAGAATAACCC 1049

771 TGTGCTTATTTGAGAGCCAGGAGTGTGTAAGAGCTGTGTTGAAGGTTTACACAAAG 830

1050 CGTCTCTCATCGCGGAGCCGCGCTGTCGAAGACGCGCGCTGTTGAGAGGCGCTCCGCGAGC 1109

831 GATTGTGAAGAGGAGATGTCGCCAAGCTCTTACTGATGTGAGATTAATTTCTGTTGACAT 890

1110 CATGCTTCCGCGGAGCTGTCGCCAGTAACCTCTCTGAGCTCCGCTCATCGCGCTGACAT 1169

891 GGGTCCGTTAGTTCTGCTCTAAATACCGAGGAGATTTGAAGAAAGGTTGAAATCTGT 950

1170 GGGCGCGCTGTCGCGCGCGCAAGTACCGCGCGAGTTCGAGGAGCGGCTCAAGCGCT 1229

951 TTTGAAAGAGTTGAGGAGCTGAGGCAAGTCTCTTTTATTGATGAGATTCATTT 1010

1230 GCTCAAGGAGTGAAGAGGCGCGAGGGAAGGTCATTTCTTTCATCGACGAGATACACT 1289

1011 GGTTCCTGCTGCGCAAACTGAAGGCTGATGATGAGCTAATCTGTTCAAGGCCAT 1070

1290 GGTCTGCGCGCGGAGGAGGAGGCTTCAAGGAGCGCGCAACCTGTTCAAGCCAAT 1349

1071 GTTAGCTAGAGGCGAGCTTGTGATGCTGCTCAAGCTTGAAGAAATACAGGAATA 1130

1350 GCTGCGAGGAGGAGCTGAGTGCATCGCGCGCACCACTGAGGAGTACCGCAAGTA 1409

1131 TGTGAGAAAGATGCTGCTTTCGAGAGGAGTTCACAAAGTCTATGTCGGAGCGCAAG 1190

1410 GGTGAGAGGAGACGAGCGTTCGAGCGCGGTTCCACAGTGTCTGTCGGAGCGGAG 1469

1191 TGTGCTGACACCAATTTAGTATCTTTAGAGGACTCAAGGAGAGTATGAGGAGCATATGG 1250

1470 GGTGCGGACACCGCTGAGCATTTCTGAGGGGCTCAAGGAGAGTACGAGGCGCACCATGG 1529

1251 TGTGGAATCAAGACAGAGCTCTTAAATGCTGCTCAGCTGCTGCTCTGTTACATAAC 1310

1530 GGTGAGATCAGGACCGCGCTCTGCTGCGGCGACAGCTCTCCGCGAGGTACATCAT 1589

1311 TGTGCGCATTTACCGGATTAAGCAATTTGTTGATGAGGCTTGTGCGAATGTGAG 1370

1590 GGTGCGGACCTGCTGACAAAGCCATAGACCTGTTGAGCGAGGCTTCCGCAATGTGAG 1649

1371 AGTCCAGCTTGAATGTAACCTGAAAGATTTGATACTTTGAAGGAGAGGATGAGCT 1430

1650 GGTGAGCTGACAGCCAGCGGAGAGATTTGATACTTTGAGAGGAGAGGATTCAGCT 1709

1431 GGAATTTCAATCTTCAAGCTTGAAGGAGAGGAGGAGTAAAGCCAGCAAGCTCGACTTAT 1490

1710 TGAAGTTGAGTTCACGCGCTCGAGAGGAGGAGGAGCAAGGCCAGCAAGCCCGGCTGAT 1769

1491 AGAGGTGC----- 1498

1770 TGAAGTGGTGTCTCGAGCACTGAATTTTCTCAAAAGTCTCTTGTCTGTGATGTTCTCT 1829

1499 -----GGAAGAG 1506

1830 GTTTGAACTTTTGATCTAACGCTGCTTAACCTGTGGCATTTGTCAGGTTCAGAGGAA 1889

1507 CTTGATGACTGAGAGCAAGCTTCAAGCTCTCAGATGAATACAGAAAGGAGAAAGAG 1566

1890 TTGAGCAGTCTGAGGAGCAAGCTGAGGCGCTTACCATGATGAAGTACAGAGGAGAGAG 1949

1567 AGAATTGATGAGATTCGAAGGCTTAAACAGAAAGAGAGAGGCTCATGTTTCTTTTCAG 1626

Transgenic plants with improved heat stress tolerance, useful for producing animal feed, oil and synthetic products -

Claim 4; Page -; 91pp; English.

The invention relates to a transgenic plant, comprising a genetic construct comprising a promoter operatively linked to a nucleic acid sequence (AA16057-AA16084) encoding a plant Heat Shock Protein (HSP) family amino acid sequence (AA16057-AA16084). The transgenic plant has increased stress tolerance, especially to heat. The plant is a cereal, grass, ornamental plant, crop plant, food plant, oil-producing plant, a synthetic product-producing plant, an environmental waste absorbing plant, an alcohol plant, a medicinal plant, a recreational plant and/or an animal feed plant. In particular, the transgenic plant is cotton, canola, soybean, corn, wheat, tobacco, sorghum, potato, tomato and/or Arabidopsis thaliana. The plants may be used to produce animal feed, alcohol, crop, oil, medicine or a synthetic product.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained from GenBank using the Accession Number reference provided in the specification.

Sequence 3942 BP; 834 A; 1111 C; 1236 G; 760 T; 1 other;

Query Match 26.0%; Score 807; DB 22; Length 3942;

Best Local Similarity 58.5%; Pred. No. 6.6e-212;

Matches 1918; Conservative 0; Mismatches 845; Indels 514; Gaps 9;

117 CGAGTTCTCATATATCGTTTAAAGGATTAACAAAGCTAATCGAAGATCAATCCAGAGAA 176
390 CCGCTTGTGTAGTATGTGAGGACTGAGGACCGAGAGCAGCCATGAATCCGACAA 449
177 ATTAC 236
450 CTTAC 509
237 ACATGCTCAATTCACCTCTTTGATTTAGTGTGTTTGTATCTTGATCCACCGGTAT 296
510 CCACGGCGAGCTCACGCGCTGACCTGCGCGCAGTGTGCTGCGACACAGGGCGCAT 569
297 ATTCTCTCAAGCAATCTTAGTCCGGTGGCAGAGACGAGC-----TCAATCTGCTGA 350
570 CTTGCGGCGAGGCCATCACGGGGGCGTGGGGGGGCGAGCGCGCGGGGACTCGTTTGA 629
351 AAGAGTGATCAATCAAGCCCTTGAAGAGCTTCTTCAACATCTCTCCACCTGATGATAT 410
630 GCGGTGCTGAACACTCGCTCAAGAGCTGCGCTGCGAGTCCCGCGCGGACTCCGT 689
411 TCAGCGAGTCTTAGTCTTATTAAGTCAATCGTGTGCTCAAGCTGCTCAGAGTCAG 470
690 TCCGGCGCTTACGCGCGTGTATCAAGGTCAATCCGCGCGGCGCAGTCCGCGCGAAGAAC 749
471 AGGTGATCTATTGCTGCTGACCAAGTGTATGAGTGTCTTTGAAGATCTCAAAAT 530
750 CCGGGAATCGCACTCGCTGCTGACCAAGCTGCTGCTGCGCTGCTGAGGACTCGAGAT 809
531 CAGGGAATTTGTTGAACGAAGTCCGTTAGCAGCGCGGAGGTTAAAGTCTGAGTTGAGAA 590
810 CTCGACTGCTCAAGGAGCGCGGTGTCGCGGCGCGGTGCGCGCGCGGCTTGAGAA 869
591 GCTTCGCGGGAAGAGGAGAAAGTTGAGAGTGTCTTACGGGACACAAATTTCAAGC 650
870 GCTCCGCGGGGAGGCGCGCGCTGAGTCCGCGTCCGCGGGGATACCAACTTCCAGGC 929
651 TTTAAAGACTTATGGAAGATTTGTTGAGCAGCAGGAGGAGCTTGATCTCTGTGATTGG 710
930 GCTCAAGACATACCGCGGGACCTCTGTCGAGAGCGCGGAGAGCTTGACCCCTCATCGG 989
711 TCGTGTAGGAGATTGAAGAGTCTGAGAGATTTCTTTTCGAGGAGAACGAGAACCAATCC 770
990 CCGCGACGAGGAGATCCGCGCTGCTGCGCATTTCTCTCGCGCGCAGCAGCAATACCC 1049
771 TGTGCTTATTGAGAGCGAGGAGTTGGTAAACAGCTGTGTTGAAGTTTACGACAAAG 830

1050 CGTCTCATCGCGAGCCCGCGTGTGCAAGACCGCGCTGTGAGGCGCTCGCGAGCG 1109
831 GATTGTGAAGAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAAATTTGTTGACAT 890
1110 CATGTTTCGGGCGAGCGTGTCCAGTACCTCTCCGAGCTCGGCTCATCGGCTCGACAT 1169
891 GGGTCCGTTAGTTGCTGCTAAATAACGAGAGAGTTTGAAGAAAGGTTGAATCTGT 950
1170 GGGCGCGCTCGTCCGCGCGCCCAAGTACCGCGCGAGTTTCGAGGAGCGGCTCAAGCGCGT 1229
951 TTTCAAGAAAGTTGAGGACCTGAAAGCAAGTGTCTCTTATTGATGAGATTCATTT 1010
1230 GCTCAAGAGGTTGAAGAGCGCGGAGGAGGTCATCTCTTTCATCGACGAGATACACT 1289
1011 GGTCTTGTGCTGGCAAAACTGAAGGTCGATGAGTGCAGCTAATCTGTTCAAGGCCAT 1070
1290 CGTCTGGCGCGGCGAGGAGGTTTCATGAGCGCGCCCACTGTTCAAGCCAT 1349
1071 GTTAGCTAGAGGCGAGCTTCGATGCTTGTGTACAACTTTGAAGAAATACAGAAATA 1130
1350 GCTGCGAGGCGACAGCTCAGGTGCTATCGCGCGCCACCCAGCTCGAGGAGTACCGCAAGTA 1409
1131 TGTGAGAAAGATGCTGCTTTCGAGAGAGGTTCCACAAAGTCTATGTTCCGAGGCAAG 1190
1410 CGTGAGAGAGACACGCGTTCGAGCGCGGTTCCAGCAGGTGTTGTCGCGGAGCGAG 1469
1191 TGTGCTGACACCACTTATGATCTCTTAGAGGACTCAAGGAGAGTATGAGGAGCATCATGG 1250
1470 CGTCCCGCAGACCTGCTGAGGCTCTGAGGCGCTCAAGGAGAGTACGAGGCGCACCATGG 1529
1251 TGTGCAATCAAGACAGAGCTCTTATAAATGTGCTGAGTGTCTGCTGTTACATAAC 1310
1530 CGTGAGGATCCAGGACCGCGCTGCTGCTGCGCGCACAGCTCTCCGCGAGGTACATCAT 1589
1311 TGTGCGCATTTACCGGATAAGCAATTTGTTGATGAGGCTGTCGATGTGAG 1370
1590 GGGTGGGACCTGCTGACAAAGCCATAGACCTGTTGAGAGGCGCTCGCCCATGTGAG 1649
1371 AGTCAGCTTATAGTCAACCTGAAGAGATTTGATTAACCTTGAAGAGAGAGATGACGCT 1430
1650 GGTGAGCTCGACAGCGCGGAGAGATTTGATACTCGAGAGGAGAGATCCAGCT 1709
1431 GGAAATTTGAATTTCAAGCTTGAAGAGAGAGATTAAGCAGGAGAGCTGACTTAT 1490
1710 TGAGTTGAGCTTCCAGCGCTTCGAGAGAGAGAGAGACAGGCCAGCAAGCGCGCTGAT 1769
1491 AGAGTGC----- 1498
1770 TGAGTGTGCTTCGAGCACTGAATTTCTCTCAACAGTCTCTTGTCTGATGTTCT 1829
1499 -----GGAAAGAG 1506
1830 GTTTTGAATTTTGATCTAAGCTCGCTTAACTGTGCGCATTTGTTCAAGGTCAGGAGGAA 1889
1507 CTTGATGCTGAGAGCAAGCTTCAGCTCTCAGGATGAATAACAGAGAGAGAGAG 1566
1890 TTGAGCATCTGAGGAGCAAGCTCGCGCGCTGACCATGAAGTACAGAGAGAGAGAG 1949
1567 AGAATTTGATGAGTTTCGAAGCTTAAACAGAGAGAGAGAGCTCATGTTTCTTTCGAG 1626
1950 AGAATTTGATGAGTACAGAGCTGAGCGCGCGCGAGGAGCTCCAGTTCCACCTCGAG 2009
1627 GAGGCGAGAACGAGATATGACCTTGCAGAGCTGCTGATCTAAGATATGCGCGCAATTCAA 1686
2010 GAGGCGAGCGCGGATGGAACCTGCGCGCTGTGCGCGACCTCAAGTACGCGCGCGCTCCAG 2069
1687 GAGTGGATCTGCAATTTGCCAATTTGAGGAACTTTCTTCTGAGAGAGATGTCATGCTC 1746
2070 GAAATTCGACGCTGCTATCTTCAAGCTGGA-----GAGCGAAACAGGGGAGAACCTGATGCTC 2126
1747 ACAGAAACGTTGGCGCTGAACACATTCGTC----- 1777

2127 ACCGAAACCGTCGCGCCCTGAACAAATTGCAGAGTATGTTATTATTCTTGTTCACCGTCA 2186
1778 ----- 1777
2187 CAAAAATTTTGCAGAGCAAGTGCAGAAATTTGCCGATCGTCTCCTAGTAGAGTAGTCGTGC 2246
1778 -----AGTTGTGAGCCGTTTGACACAGGATTC 1805
2247 AGCGTCTGAAATGGTGTTCCTTATGGCAGGTTGAGCGGTTGACGGGATTC 2306
1806 AGTACAGAGCTTGCCCAAAATGAGAGAGAGGTTGATGTTGCTTCTGTATAGGTTGCA 1865
2307 AGTACCCGCTTGCCGAGAACGACAGAGAGAGGCTGTTGGCTCTGCTGACAGGCTTCA 2366
1866 TAGCGGTTTGGGACAGATCAAGCGTAAATGCAAGTTTCTGAGCRAATCTTAAGTTC 1925
2367 CCAGAGGTTGTCGCGCAGACAGAGGCTGTGAGCGCTGCGAGAGCGGTTCTGAGTTC 2426
1926 AAGGCGAGGACTTGTGAGGCGACAAACAGCCAACTGGATCATCTTATTCTTGGACCAAC 1985
2427 GAGGCGGCTTGTGAGGCGACAAACAGCCCACTGGCTCTCTCTCTGCTGGTCCGAC 2486
1986 TGGTGTGGCAAACTAGCTGCGCAAGGCTCTTCTGAGCAGCTGTTTCAATGATAA 2045
2487 TGGCGTGGGAAAATGAGCTGGCCCAAGGCGCTAGCCGAAACAGCTGTTGACACGAGAA 2546
2046 CCTCTAGTTCCGATTGATGTCGGAATATATGGAACAACTCTGCTCTCTGCTCAT 2105
2547 CCTCTTGTCCGATCGACATGTCGGAGTACATGGAGCAGCATCGGTTCCGCTCAT 2606
2106 TGGGCGCCACACAGGCTATGTT----- 2127
2607 CGGAGCACCACTCGGTAAAGTAGCAGAAATGATGATCTTCTGATTTTAACTCGAA 2666
2128 -----GGTCAGAG 2136
2667 TTGCGATGAATCTTGTCTGACAGTGGCGGACCTGTGACAGCTACGTCGGCCATGAA 2726
2137 GAAGTGAGCAACTAACTAGGCTGTGAGGAGCGGACCTATTGTGTGATCTTCTTGTAT 2196
2727 GAGGTGGGAGCTGATGTAACAGTGAAGGAGGCGGTACAGCGTATCTTCTGAC 2786
2197 GAAGTGAGAGGCTCATGTTGTCTTCAACTCTGCTCCTCAAGTTTGGATGATGTT 2256
2787 GAGTTCGAGAGGCGCCATGTCGCGTGTTCACACCTGCTTCCAGGTCCTCGAGCGGC 2846
2257 CGATTGACAGCGGCGAAGGAGAGAGTGAATTTAGAGAACTCGGTGATTAATGACACA 2316
2847 AGGCTGACGGATGGGCAAGGAGGAGCGTGACTTCAGGAACACCGTGTATCATGACA 2906
2317 TCAAACTTGTGCTGAACACTCTTTCAGGCGTAACTGGGAAAGTAACTGGAAGT 2376
2907 TCGAACCTCGGCGGAGCACCTCTCTGCTGGATGTTGGCAGAACTCCATGAAGTTC 2966
2377 GCCCGGAGTGTGTGATCGGG----- 2398
2967 GCTCGGATCTGCTCATGACGAGGATGATGATGCGATTAAGCTTCTGAATCGTCTTGA 3026
2399 -----AGGTGAGAA 2408
3027 CATTTGTTCTCTGACGCGTGAATGCTCAATCGGATGCTTATGCTGTGTGAGGTTGAGG 3086
2409 ACATTTACAGACAGAGCTTTGAACAGGCTTGAAGAGTGTGTGTTTGAACCCCTTTC 2468
3087 GCATTTCCGCTGAGCTGTGAACCGTCTCGACGAGATCGTGAATCTGATCTCTGTC 3146
2469 ACATGACAGTTGAGGAAAGTAGTCTCGGCTTCAATGAAGACGTTGCTGTCGGCTTC 2528
3147 CCACGAGCAGCTGAGGAGGTCGCTCGCTTCAGATGAAGATGTTGGCGCTCTCTTC 3206
2529 TGAAGAGAGGTTGTTTGGAGTCACTGATGCTGCTTGAACATATATCTTGGCAGAGAG 2588
3207 CGAAAGGGGATCGCTCTGGCTGTGACCGGCGCATTTGGACATCATCTTGTCTCTC 3266

QY 2589 TTATGACCC----- 2597
DB 3267 TTACGATCGGTATGTGACCATCCATGATTTGATCCATCTGATTCGTCGGTGACACCTG 3326
QY 2598 -----GGTGTATGTCGTAGC 2614
DB 3327 ATGGTCTGACTCTCTTATCTTCTTGTGTGCTTCAACAAACAGGTGTATGCGCGCGC 3386
QY 2615 CTATAGAGATGAGTGAAGAGAGGTTGTCACAGAACTGTCAAGATGTTGTGCGTG 2674
DB 3387 CAATCAGAGGTGATGAGAGAGGTTGTCACAGCTGTGCAAGATGCTGATCCAGG 3446
QY 2675 AGGAAATCGATGAAATCTCACTGTTTACATAGATG-----CAGGCGCTGCTGATCTT 2728
DB 3447 AGGATCGAGAGAACTGACGGTCTACATCGACCGCGCGGCAAGGACGAGCTGG 3506
QY 2729 TGTACCGGTAGA-----AAGTGAAGTCTAGTGAGGCTTCAACAGGCAAGAGTCAATG 2785
DB 3507 TCTACAGGTTGACCGGAGCGCGCTCTGTTGAACGCTGAGACGGGATGAAGTCGACA 3566
QY 2786 TGCTGATTCATATCTTAACGGGCCAAAGAGAGATGATGACAGCTCAGGCGGTGAAGAAGA 2845
DB 3567 TCTGATCCAGTCCCAACAGCTCCACAGAGGAGCGCTGCGGAGGCCCTCAGAGAAGA 3626
QY 2846 TGAGGATCGAGAAATAGAAGATGACGATAATGAGGA 2882
DB 3627 TGAGGATCATGAGGAGGACGAGGACGGCATGAGGA 3663

RESULT 14

AA13093
ID AA13093 standard; DNA; 6491 BP.
XX
AC AA13093;
XX DT 19-MAR-1999 (first entry)
XX Enterococcus faecalis genome contig SEQ ID NO:156.
DE Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX Enterococcus faecalis.
OS WO9850555-A2.
XX PN 12-NOV-1998.
XX PF 04-MAY-1998; 98WO-US08985.
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046655.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Barash SC, Dillion PJ, Kunsch CA;
XX WPI; 1999-045171/04.

XX New isolated Enterococcus faecalis polymucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX Claim 1; Page 885-888; 2084pp; English.
XX A computer readable medium has been developed which has recorded on it
XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX AA12938 to AA13919 represent these nucleotide sequences which are
XX primary nucleotide sequences, also known as contigs. The computer-based
XX system can identify fragments of the Enterococcus faecalis genome with

commercial importance. The products can be used to detect the presence of *Enterococcus faecalis* in samples. They can also be used for diagnosing *Enterococcal* infection in an animal and monitoring progression of disease, and for identifying agents which can be used to modulate the growth or pathogenicity of *Enterococcus faecalis*, or another related organism, *in vivo* or *in vitro*. In particular the polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences can be used in vaccines to prevent or attenuate an *Enterococcal* infection.

Sequence 6491 BP; 2164 A; 1114 C; 1351 G; 1844 T; 18 other;

Query Match 21.7%; Score 672.8; DB 20; Length 6491;
 Best Local Similarity 59.1%; Pred. No. 1e-174;
 Matches 1180; Conservative 5; Mismatches 807; Indels 6; Gaps 2;

683 AAGCAGGAGCTTGAATCTGATGTTGGTGTGATGAGGAGATTAGAAAGATCGTGAGGA 742
 546 AAGCAGGAGGAAACAAGATCAATCATTTGGTGTGACGAGAAATTCGTGACGTTATTCGGA 605
 743 TTCTTTTCGAGGAGACGAGACAACTCTGTGCTTATTGGAGAGCCAGGATGTTGTAATA 802
 606 TTTTATCAAGAAATACTAAATAATCCAGTCTTAATTGGTGAACAGGATGTTGGTAATA 665
 803 CAGCTGTGTTCAAGGTTTAGCACAAGAGATTGTGAAAGGAGATGTGCCCAACAGTCTTA 862
 666 CAGCGATTGTTGAAGGATTAGCACAACGAATGTTTGGTAAGATGTTCCCGAAACTTAA 725
 863 CTGATGTGAGATTATTCGTTGGACATGGTGGTGTAGTTAGTCTGCTGCTTAATACCGAG 922
 726 AAGATAAAACCAATTTTCTTTAGATATGGGCGCTTAATTGCGGGCAAAATTCGCTG 785
 923 GAGAGTTCAAGAAAGTTGAATCTGTTTCAAGAGAGTTCAGACGCTGAAGGCAAAAG 982
 786 GCGAATTTGAAGAACCGTTAAAGCTGTTTAAAGAGTGAAGAAAGTGTATGSCAAA 845
 983 TGATTCCTTTATGATGAGATTCAATTTGGTCTTGGTCTGGCAAAACTGAAGGTCGA 1042
 846 TCATTTTATTCATTGATGAATTCATAATATCGTCGGAGCTGTGTAACAGAGGACGTA 905
 1043 TGATCAGGTTAATCTGTTCAAGCCCATGTTAGTCTAGAGGGCAGCTTCGATCATTGTTG 1102
 906 TGACGCGCGGAAATTTATTAACCAAGATGCTAGCAGCTGTGAAATTCGCAATTTAATCGGTG 965
 1103 CTACACGCTGGAAGATACAGAAATATGTTGAGAAAGATCTGCTTTTGAGAGAGGT 1162
 966 CCAACAGCTTGATGAATATCGCAATATATGAAAGAGATTAAGCTTAGAACGTCGTT 1025
 1163 TCCAAACAAGTCTATGTTGGGAGCCAAAGTGTGCTGACACCATTTAGTATCTTTAGAGGAC 1222
 1026 TCCAAAGATATTAGTCAAGAACCAACAGTAGAAGATACAAATTTCTATTTTGGCTGGCT 1085
 1223 TCAGAGAGATGAGGACATCATGTTGGATTCGGAATCCAGACAGAGCTCTTATAAATG 1282
 1086 TAAAGAACGATTGAATTTACCCAGCGGTGAATATCATGACAAATGCTTTAGTTGCTG 1145
 1283 CTGCTCAGCTGTCTGCTGTTATCAATACTGTCGCAATTTACCGGATAAGCAATTTGATT 1342
 1146 CCCTACTTTATCCGATCGCTATATTACTGATCGTTCTTACAGATAAGCGATTGACT 1205
 1343 TGGTTGATGAGGCTTTGCGAATGTGAGAGTCAGGTCAGGTTGATAGTCAACCTGAGAGATTG 1402
 1206 TAGTCGATGAAGCAAGCGCAACGATTCGAGTGAATGAATTCGATGCCAACAGAGCTCG 1265
 1403 ATACCTTTGAAGGAGAGATGCGATGGAATTCGAACTTCACGCTTTGGAAGGAGA 1462
 1266 ATCAAGTAAACAGCTCGGTAAATGACATGAGAAATGAAGAGCGGCTTTGAAAGAAAT 1325
 1463 AGGATAAGCCAGCAAGCTCGACTTATAGAGGTGCGGAAAGAGCTTTGATGACTGAGAG 1522
 1326 CAGATGATGCAATGAAGAAACGTTAGCCCAACTTGCAGAGAGATTTAGCAGATTTCGCTG 1385
 1523 ACAAGCTTCAGCCTCTCAGATGAATACAGAAAGGAGAGAAATGATGAGATTTC 1582

1386 AAGAGCCCAACTCAATGAAATGCAATGGAAACGGAACGAAAGAGTTAATGCGGTTT 1445
 1583 GAGGCTTAAACAGAAACAGAGAGAGTCTATGTTCTTTGAGGAGGAGAGCAACGAAGAT 1642
 1446 CCAATAACGTCGAGAGATGATAAGCAAAACAGCAATTAGAGATGCTGAAATAATT 1505
 1643 ATGACCTTGCAGAGCTGCTGATCTAAGATATGGCGCAATTCAGAAAGTGGAACTGCA 1702
 1506 ATGATTTAGAACGAGCTGCTGTTTTCGTCATGGAACAATTCGCAATTAGAACATGAT 1565
 1703 TTCCCAACTTGAAGAACTTCTTGAAGAGATGTGATCTCAGCAAAACGTTGGGC 1762
 1566 TGAAGAAATTTGAAGAAAGAGACGCTTAAAGATAACGCTCAAAATGGTACAAAGATCGTTA 1625
 1763 CTCAACAC---ATTGCTGAGGTTTGTGAGCGGTTGAGACAGGATTCAGTGCAGAGACTTG 1819
 1626 CTGAAACGMAATTTGCGAAGTGTGCTGCTGTTAATCTGCAATTCCTGTGACAAATTAG 1685
 1820 GCCAAATGAGAAAGGAGAGGTTGATTGGTCTTCTGATAGGTTGCAATAGCGGTTGTGG 1879
 1686 TTGAAGGCGAACGAGAAATAATTAAGAACTGAAATGAACAATTAACAACAACGTTGATTG 1745
 1880 GACAGATCAAGCGGTAATGCAAGTTCTGAGGCAATTCAGGTCAGAGGCGAGACTTG 1939
 1746 GTCAAGATGAAGCGGTCGATGCTGTAGTGATGCGGTGATTCGTTCAAGAGCAGGCTTAC 1805
 1940 GTAGGCGCACAAACAGCCTGGAATCAATCTTATTTCTTGGACCAACTGGTGTGGCAAAA 1999
 1806 AAGATCCAAATCGCCCACTCGGTTCGTTCTTCTTAGGACCAACTGGTGTGGTAATA 1865
 2000 CTGAGCTCGCAAGGCTTCTGTCGAGAGCTGTTGATGATGATGAAACCTCTTAGTTCCGA 2059
 1866 CAGAACTTGCTAAAGCTTTAGCTCAAGATTTGTTGATTTCTGAAGATCATATGTTACGGA 1925
 2060 TTGATATGTCGGAATATATGGAACACACTCTCTCTCTGCTCCTATTGGGGCACCAACAG 2119
 1926 TTGACATGATGATATCATGGAACCAATGCGGTGCTCGTTGGTTGGTGGCCCTCCAG 1985
 2120 GGTATGTTGTCACAGAGAGGTGGCAACTAACTGAGGCTGTGAGAGGCGACCTTATT 2179
 1986 GCTATGTTGTTGAGAGAGTGGCCAGTTTAAAGGAGCTGTTGCTGCAACCTTATA 2045
 2180 GTGTCTACTCTTTGATGAGAGGAGGCTCATGTTGCTGCTCTTCAACACTCTGCTCC 2239
 2046 CAATGTCTTATTAGACGAATTTGAAAGAGCGCACCAAGATGCTTTAATATCTTATTAC 2105
 2240 AAGTTTGGATGATGCTGATTCGACAGCGGCAAGCAGACAGTCGATTTTCAGGAAT 2299
 2106 AAGTTTGTAGATGATGCTGTTGACAGATTCCTAAGTCCGGTCTGTTGATTTTAAATA 2165
 2300 CGGTGATAATCATGACATCAAAACCTTGGTCTGCAACACCTCTTGCAGGGCTAACGGGA 2359
 2166 CAGTTTAAATTTATGACAAAGTAAACATCGCTCTCAGCTTTTATTGGAAGGTGTGACACTG 2225
 2360 AAGTAAACAAT---GGAAGTGGCCCGGAGCTGTGTGATGCGGAGGTGAGGAAACACTTCA 2416
 2226 AAGCAACAATTTCCAGAGAAAGTTGAAATCAAGATCAATGAATATTTTAAAGGACACTTTA 2285
 2417 GACCAGAGCTCTTGAACAGGCTTGACGAGATTTGGTGTTCGACCCCTTTTCACATGACC 2476
 2286 AACAGAGGTTCTTAAACAGAAATGATGATCCATTTTATTTCACCAATTAAAGCTTGGATA 2345
 2477 AGTTGAGAAAGTAGCTCGGCTTCAATGAAAGACGTTGCTGCCGGCTTGTGGAAGAG 2536
 2346 ATGTGAAGGAATCATTTGTTAAATGACCGCACAACTTGTCTCATCTGTTTAGAGCAACAAG 2405
 2537 GAGTTGCTTTGGCAGTCACTGATGCTCTTTGGCTATATCTTGGCAGAGAGTTATGACC 2596
 2406 AAATTTGTTGGAATCACTGATGAAGCTAAACATGATTAAGATTTGCTGAAATTTGGTATGAGC 2465
 2597 CGGTGATGCTGCTAGGCTTATAGGAGATGGAATGAGAGAGGAGGTTGTCAGCAACACTGT 2656

2466 CAGCTTATGAGAGCTGCTCGTTAAACAGTTTCATTACTCGGAGTGGAAACACCACTTAG 2525

2657 CAAAGATGTTGTGCTG 2674

2526 CAAAAGAAATGTTCTG 2543

LT 15

8988

ABS98888 standard; DNA; 6491 BP.

ABS98888;

18-DEC-2002 (first entry)

Enterococcus faecalis contig sequence #156.

Computer readable medium; Enterococcus faecalis; microbe; growth; pathogenicity; vaccine; resistance; Enterococcal infection; commercial; therapeutic; industrial; fermenting; sugar source; metabolite; vaccine; biotech technology; antibacterial; modulator of nucleic acid expression; contig; ds.

Enterococcus faecalis.

US2002120116-A1.

29-AUG-2002.

04-MAY-1998; 98US-0070927.

04-MAY-1998; 98US-0070927.

(KUNS/) KUNSCH C A.

(DILL/) DILLON P J.

(BARA/) BARASH S.

Kunsch CA, Dillon PJ, Barash S;

WPI; 2002-750065/81.

Computer readable medium having recorded on it a Enterococcus faecalis nucleotide sequence useful for detecting diseases related to Enterococcus infections in animals

Claim 1; Page -; 119pp; English.

The present invention relates to a new computer readable medium with an Enterococcus faecalis nucleotide sequence. The invention is useful to diagnose the presence of E.faecalis in a sample or determining the presence of a specific microbe in a sample. The invention is also useful for modulating the growth or pathogenicity of E.faecalis, in a vaccine to confer resistance to Enterococcal infection, for commercial, therapeutic and industrial purposes, and for fermenting a particular sugar source or to produce a particular metabolite. The invention is useful for detecting diseases related to Enterococcus infections in animals, and for detecting E.faecalis using biotech technology. The present nucleic acid sequence represents an Enterococcus faecalis contig DNA sequence of the invention.

Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from USPTO at <http://seqdata.uspto.gov>.

Sequence 6491 BP; 2164 A; 1114 C; 1351 G; 1844 T; 18 other;

Very Match 21.78; Score 672.8; DB 24; Length 6491;
Best Local Similarity 59.18; Pred. No. 1e-174;
Matches 1180; Conservative 5; Mismatches 807; Indels 6; Gaps 2;

683 AAGCAGGAGAGCTTGATCCCTGTTGATGCTGATGAGGAGATTAGAAGAGTCTGTGAGGA 742

546 AAGCAGGAGAGAGATCCATCATCTGTTGCTGACGAGAGAAATTCGTGACGTTATTCGGA 605

QY 743 TTCTTTGAGAGAGAACGAGAACAAATCCTGTGCTATTATGAGAGCCAGAGTGGTAAAA 802
DB 606 TTTTATCAAGAAAACTAAAAATAATCCAGTCTTAATTTGTTGAACAGGTGTTGGTAAAA 665
QY 803 CAGCTGTGTTGAAGGTTTAGCCAAAGGATTGTGAAGGAGATGTCGCCAACAGTCTTA 862
DB 666 CAGCGATTGTTGAAGGATTAGCACACGAATCGTTCGTAAGATGTTCCCGAAACTTAA 725
QY 863 CTGATGTGAGATTAAATTTCTGTCGACATGGTGGTGTGTTAGTTGCTGTGCTAAATACCGAG 922
DB 726 AAGATAAAACCAATTTTCTTTAGATATGGCGCTTAATTTGCGGAGCAAAATTCGCTG 785
QY 923 CAGAGTTGAAAGAGGTTGAAATCTGTTTGAAGAGATTGAGGACGCTGAAGGCAAG 982
DB 786 CGGAATTTGAAGACGGTTAAAGCTGTTTTAAAGAGTGAAGAAAAAGTATGCGCAAAA 845
QY 983 TGATTTCTTTTATGATGAGATTCAATTTGTTGTTGCTGTGCGCAAACTGAAGGTCGA 1042
DB 846 TCAATTTTATTCATTTGATGAATTCATTAATTCGTCGGAGCTGTGTAAGACAGAGGCAAGTA 905
QY 1043 TGGATGAGCTTAATCTGTTCAAGCCCATGTTAGCTAGAGGCGAGCTTCGATGCAATTCGTTG 1102
DB 906 TGGACGCGGAAATTTTATTAACCCGATGCTAGCACGTGGTGAATTCATTTAATTCGTTG 965
QY 1103 CTACAAAGCTTGAAGAGATACAGGAAATATGTTGAGAAAGATGCTGCTTTGAGAGAGGT 1162
DB 966 CCACAAAGCTTTGATGAATATCGCCAATATATGGAAGAAAGATTAAGAGCCTTAGAACGCTGTT 1025
QY 1163 TCCAAAGCTATGTTGCGGAGCCAAAGTGTGCTGACACCATTTAGTATCTCTTAGAGAGAC 1222
DB 1026 TCCAAAGAGTATTAGTCAAGAACCAACAGTAGAAGATACAAATTTCTTTTGTGCTGGCT 1085
QY 1223 TCAAGGAGAGATATGAGGAGACATCATGTTGTCGAATCCAAAGACAGAGCTTTATAAATG 1282
DB 1086 TAAAGAAACGATTTGAAATTCACCAAGCGGTGAATATTCATGACAAATGCTTTAGTTGCTG 1145
QY 1283 CTGCTCAGCTGTCTGCTGTTACATAACTGTCGGCATTTACCGGATTAAGCAATTCGATT 1342
DB 1146 CCGTACTTTTATCCGATCGCTATATTAATGATCGTTTCTTACCAGATAAGCGGATTGACT 1205
QY 1343 TGGTTGATGAGCTTTGTGCGAATCTGAGAGTCCAGCTTGATAGTCAACCTGAAGAGATTG 1402
DB 1206 TAGTCGATGAAGCAAGCGCAAGATTCGAGTGGAAATGAATTCGATGCCACAGAGCTCG 1265
QY 1403 ATAACCTTTGAAGGAGAGGATGAGCTGGAAATTTGAACCTTCACGCTTGAAGAGGAGA 1462
DB 1266 ATCAAGTAAACGCTCGGTTAATGCMACVAGAAATTTGAAGAGCGGCTTGAAAAAGAAAT 1325
QY 1463 AGGATAAGCCAGCAAGCTCGACTTATAGAGTGGCGAAGAGCTTTGATGACCTGAGAG 1522
DB 1326 CAGATGATGCAAGTAAAGAAACCGGTTAGCCAACTTGAAGAGAAATTTAGCAGATTGCGTG 1385
QY 1523 ACAAGCTTTCAGCTCTCAGATGAAATACAGAAAGGAGAAAGAGAAATTTGATGAGATTC 1582
DB 1386 AAGAGCCAACTCAATGAAATGCAATGGGAAACGAAAGAGAGATTAAATCGGTTT 1445
QY 1583 GAAGCTTTAAACAGAAAGAGAGAGCTCATGTTTCTTTCGAGAGGCGAGAACAGAT 1642
DB 1446 CCAATAAACGCTGAGAGATTGATAAAGCAAAACCAAGAAATTTAGAAGATGCTGAAATAAT 1505
QY 1643 ATGACCTTGAAGAGCTGCTGATCTTAGATATGGCAATTCAGAAAGTGGAAATCTCAA 1702
DB 1506 ATGATTTAAGACGAGCTGCTGTTTGGGTGATGGAACAATTTCCCAATTTAGACATGAAT 1565
QY 1703 TTGCCCAACTTGAAGAACTTCTTCTGAAGAGAAATGTGATGCTCACAGAAACCTTGGGC 1762
DB 1566 TGAAGAGATTGGAAGAAAGAACGCTTAAGATTAAGTCAAAATGTTACAGAAATCGGTTA 1625
QY 1763 CTGAACAC ---ATTGCTCAGGTTGTGAGCCCTTGGACAGGAAATTCAGTGAACGAGATTG 1819
DB 1626 CTGAACACGAAATTTGCGCAAGTGGTGGTCTGTTTAACTGGCATTTCTGTGACAAAAATTAG 1685
QY 1820 GCCAAATGAGAGAGAGGTTGATTGCTCTTGTGATAGTTGCTGCATAGCGGTTGTGG 1879

1686 TTGAAGGCGACAGAGAAAATTAATGAACACTGAATGAACATTACACAAACGCTGTGATTG 1745
1880 GACAGAAATCAAGCGTAAATGAGTTCTGAGGCAATCTAAGGTCAGAGGCGAGACTTG 1939
1746 GTCAAGATGAAGCGCTGAGTGTGATGATGAGCGGTGATCGTTCAAGAGCAGGCTTAC 1805
1940 GTAGGCGACACAGAGCAACTGGATCATTTCTATTCCTTTGGACCAACTGGTGTGGCAAAA 1999
1806 AAGATCCAAATCGCCCACTCGGTCGTTCTCTAGGACCAACTGGTGTGGTAAAA 1865
2000 CTGAGCTCGCCAAAGCTCTGCTGAGCAGCTGTTGATGATGAAGAAACCTCTTAGTTGGA 2059
1866 CAGAAGCTTGTCTAAAGCTTAGCTGAAGATTGTTTGAATCTGAAGATCATATGGTACGGA 1925
2060 TTGATATGTCGGAATATATGGAACAACTCTGCTCTCGCTCATTTGGGGCACCACCAG 2119
1926 TTGACATGATGATGATACATGGAAGAAACATCGCTGCTCGTTGGTGGTGGCTCCAG 1985
2120 GGTATGTTGGTCACGAGGAGGTGACAACTAACTGAGGCTGTGAGGAGGCGACCTTATT 2179
1986 GCTATGKTGTTTGAAGAAGGTGGCCAGTTAACGGAAGCTGTCGTCGAAACCTTATA 2045
2180 GTGTCATACCTTTGATGAGTGGAGAGGCTCATGTTGCTGCTCTTCAACACTCTGCTCC 2239
2046 CAATTGCTCTTATAGACGAATTTGAAAAGCGCACCCAGATGCTCTTAAATATCTTATTAC 2105
2240 AAGTTTGGATGATGCTGATTTGACAGCGGAGGAGGAGACAGTGGATTTTCAGGAAC 2299
2106 AAGTTTGGATGATGCTGATTTGACAGATTTCTAAGGTCGGTCTGTTGATTTAAATA 2165
2300 CGGTGATATCATGACATCAAACTTGTGCTGGAACACCTCTTGCAGGGCTAACTGGGA 2359
2166 CAGTTTAAATATGACAGTAACTACGCTCTCAGCTTTTATTTGGAAGGTGTGACACCTG 2225
2360 AAGTAACAAAT---GGAAGTGGCCCGGACTGTGTGATCGGGAGGTGAGGAAACACTTCA 2416
2226 AAGGAACAAATCCAGAAGAGTTGAAAATCAAGTCATGAATTTCTTAAGGACACTTTA 2285
2417 GACCAAGCTCTTGAACAGGCTTGACGAGATTGTGTGTTTCGACCCCTTTTCACATGACC 2476
2286 ACCAGAGTTCTTAACAGAAATGATGATACCATTTTATTTACACCAATTAAGCTTGATA 2345
2477 AGTTGAGGAAAGTAGCTCGGCTTCAAAATGAAGAGAGCTGTGCTCGGCTTGTGAAAGAG 2536
2346 ATGTGAAGGAATCATTTGTAATAATGACCGCACAACTTGCTCATCGTTTAGAGCAACAAG 2405
2537 GAGTTGCTTTGGCAGTCACTGATGCTGCTTTGGACTATATCTTTGGCAGAGACTTATGACC 2596
2406 AAATTGTTTGGAAATCACTGATGAAGCTAAACATGGATTCCTGAAATGGGTATGACC 2465
2597 CGGTGTATGCTAGGCTTATAGGAGATGGATGGAGAGAGGTGGTGAAGAACTGT 2656
2466 CAGCTTATGAGCTCGTCCGTTAAACGTTTCATTACTCGCGAAGTCGAAACACCATTTAG 2525
2657 CAAAGATGTTGTGGTG 2674
2526 CAAAGAAATTTGTTCTG 2543

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

nucleic - nucleic search, using sw model

on: February 12, 2004, 17:26:16 ; Search time 5756 Seconds
(without alignments)

13110.736 Million cell updates/sec

le: US-09-812-350-30

fect score: 3105

quence: 1 aagttatcaatttaccac.....tctttgggggtttttcata 3105

ring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

rchd: 22781392 seqs, 12152238056 residues

al number of hits satisfying chosen parameters: 45562784

imum DB seq length: 0

imum DB seq length: 2000000000

-processing: Minimum Match 0%

Minimum Match 100%

Listing first 45 summaries

base :

EST:*

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estro:**

7: em_estro:**

8: em_hcc:**

9: gb_est1:**

10: gb_est2:**

11: gb_hcc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: em_gss_num:**

18: em_gss_inv:**

19: em_gss_pln:**

20: em_gss_vrt:**

21: em_gss_fun:**

22: em_gss_man:**

23: em_gss_mus:**

24: em_gss_pro:**

25: em_gss_rod:**

26: em_gss_phg:**

27: em_gss_vrl:**

28: gb_gss1:**

29: gb_gss2:**

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ult NO.	Score	Query Match	Length	ID	Description
1	1268.6	40.9	3163	11	AY109444 Zea mays
2	605.4	19.5	633	9	AV826643
3	603.6	19.4	1015	28	AF106743
4	560	18.0	641	9	AU236881

ALIGNMENTS

RESULT 1
AY109444
LOCUS Zea mays CL1518_1 mRNA sequence.
DEFINITION Zea mays
ACCESSION AY109444
VERSION AY109444.1 GI:21213165
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3163)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 3163)
AUTHORS Coe,S.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR, the www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the

CB894504 EST647296
BH250711 BOGAQ17TF
BH465756 BOGEI04TF
CB894478 EST647270
CB349719 CAB28G000
BG648385 EST510004
CB893989 EST646781
BH603944 BOGTL30TF
BG648376 EST646168
BG648761 EST510180
CB893763 EST646555
CB610803 ALBEDO000
BE642455 Cr12.6.B1
BG647349 EST508968
BO933972 QGF5P11.Y
AI988968 701547761
BQ869697 QG6P04.Y
BQ862449 QGC21B24.
AV795955 AV795955
BH423861 BOGKX57TR
BH492668 BOGWA22TF
BH486893 BOHSJ43TF
BF480649 LO-2608T3
BH617393 SALK_0364
CB894119 EST646911
AU227885 AU227885
BQ839239 WHE4163.H
RU892034 P058E06.P
CB894098 EST646890
BM437943 VVA028B02
BZ504339 BONQD01TR
BZ963117 PUGFU64TD
BG649062 EST510681
BJ291097 BJ291097
BZ270785 NF002A06F
BZ444584 BONQMB7TR
AW257917 687064B03
BZ292649 SALK_1249
BE515416 WHE0605.C
CB915944 VVD104G07
BF632285 NF017C06D

846 14 CB894504
818 28 BH250711
863 28 BH465756
837 14 CB894478
843 14 CB349719
736 10 BG648385
812 14 CB893989
750 28 BH603944
783 14 CB893763
818 10 BG648761
804 14 CB893763
722 14 CB610803
783 10 BE642455
747 10 BG647349
727 13 BO933972
467 9 AI988968
686 13 BQ869697
729 13 BQ862449
423 9 AV795955
689 28 BH423861
725 28 BH492668
559 28 BH486893
703 10 BF480649
448 28 BH617393
598 14 CB894119
447 9 AU227885
740 13 BQ839239
621 13 RU892034
610 14 CB894098
706 12 BM437943
551 29 BZ504339
709 29 BZ963117
598 10 BG649062
617 12 BJ291097
618 12 BZ270785
815 29 BZ444584
625 9 AW257917
398 29 BZ292649
629 10 BE515416
591 14 CB915944
567 10 BF632285

5 543.4 17.5
6 542 17.5
7 541.4 17.4
8 529 17.0
9 506.4 16.3
10 505.8 16.3
11 502.4 16.2
12 494 15.9
13 490.2 15.8
14 483 15.6
15 481.2 15.5
16 470.2 15.1
17 462.8 14.9
18 456 14.7
19 450.6 14.5
20 435 14.0
21 433.8 14.0
22 423.8 13.6
23 423 13.6
24 423 13.6
25 417.2 13.4
26 403.6 13.0
27 399.8 12.9
28 398.4 12.8
29 391.6 12.6
30 387.8 12.5
31 384.6 12.4
32 384.4 12.4
33 377.2 12.1
34 372.2 12.0
35 367.6 11.8
36 364.2 11.7
37 363.2 11.6
38 361.6 11.6
39 357.4 11.5
40 357 11.5
41 356.8 11.5
42 353.4 11.4
43 353.2 11.4
44 348.8 11.2
45 346 11.1

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.

FEATURES
source
1. 3163
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:630374"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration with the overgo addressing of BACs in conjunction with the Maize Mapping Project"

E COUNT 626 a 846 c 942 g 490 t 259 others

GIN

Query Match 40.9%; Score 1268.6; DB 11; Length 3163;

Best Local Similarity 63.8%; Pred. No. 1.4e-279; Matches 1782; Conservative 0; Mismatches 992; Indels 18; Gaps 4;

141 GGATTACAAAAGCTAATCGAAGATGAATCCAGAGAAATTCACACAGACAAACGAGAC 200
132 GGAATGACACAGAGACGAGCATGAATCCGGAACACTTACCCACAGACGACGAGGC 191
201 AATTGCTACAGCTCATGAGTAGCTGTGAATGCGAGCATGCTCAATTCATCTTTGCA 260
192 GATCGTNNNNCCGACGAGATGCGGTGGAGCGCGCCACGCGAGCTCAGCGCGTGA 251
261 TTTAGCTGGTGTGTGATCTGTATCCACCGGTATATTTCTCAAGCAATCTTAGTGC 320
252 CCGTCCCGAGTGTGGTGGCGAAGAGGGGGGATCTCTGCGGAGGCGCATCCNNNNNN 311
321 CGGTGGCGAGAACGAGCTCAATCTGCT-----GAAAGAGTGATCAATCAAGCTTTGAA 374
312 NNN 371
375 GAAGTTCCTTCAAAATCTCTCCACCTGATGATATTCACGCGAGTCTTAGCTTATTA 434
372 GAAAGTGGCTCGAGTCCCGCCCGCGAGCTCCGCTCCCGCGCTCCACGCGCTGATCAA 431
435 GGTCAATCGTGTCTCAAGCTGTCAGAGTCAGAGTGATCTACTATTGGCTGTGA 494
432 GGTCAATCGCGCGGCGAGTCCGCGCAGAGAAACGCGGGAGCTCGACCTCGCGCTGA 491
495 CCAGTTGATTATGGTCTTCTTGAAGATTTCTCAATCAGGGATTTGTTGAACGAGTCGG 554
492 CCAGCTGCTGCTGGCTGCTCGAGGACTCGAGATCTCCGACTGCTCAAGNNNNNN 551
555 TGTAGCGAGCGGAGGGTAAAGTCTGAGTTGAGAGCTTCGTGGGAAGAGGAAGAA 614
552 NNN 611
615 AGTTGAGAGTGTCTCAGGCGACACAAATTTTCAAGCTTTAAAGACTTATGAGAGATTT 674
612 NNNNNNNCCGGTNNNNNACCAATCTCCAGCGCTCAAGACATACGCGCGGAGCT 671
675 GGTGAGCAAGCAGGAGAGCTGTGATCTGTGATTTGGTGTGATGAGGAGATTAGAAGAGT 734
672 CGTGAAGAGCGGAGAGTGAACCCCGTCAATCGCGCGCAGGAGAGATCCGCGCGT 731
735 CGTGAAGATTTCTGAGAGAAACGAGAAACAACTCTGTGCTTATTTGAGAGCGCAGGAGT 794
732 CGTGCATCTCTCGCGCGCAGCAACAAACCCCTCTCATCGCGGAGCCCGCGGT 791
795 TGGTAAACAGCTGTGTGAGTTTACCAAAAGAGATTGGAAGAGAGATGTCGCCAA 854
792 TGGCAAGAAGCGGCGTGTGAGGGGCTCCGCGAGCGCATCGTTCGCGGAGCTGCCCG 851
855 CAGTCTTACTGATGTGAGATTAATTTCTGTTGACATGGGTGGTGTGCTGTGCTAA 914

852 TAACCTCTCTGAGCTCGCCTCATCGCCTCGACANNNNNNNNNNNNNNNNNNNNNNNNN 911
915 ATACCGAGAGAGTTTGAAGAAAGTTGAAATCTCTTTTGAAGAAAGTTGAGGACCTGA 974
912 GTACCGCGCGAGTTTCGAGAGCGGCTCAAGCCCGTGTCTCAAGGAGGTGGAAGAGCCGA 971
975 AGCAAGGTGATTCTCTTTATGATGAGATTCATTGTTCTTGTGCTGGCAAACTGA 1034
972 GGGGAGGTCAATCTCTTCATCGCAGATACACTCTGTCNNNNNNNNNNNNNNNNNNNNNN 1031
1035 AGGGTCGATGATGAGCTTAATCTCTTCAAGCCCATGTTAGCTAGAGGCGAGCTTCGATG 1094
1032 NNNNNCCATGAGCGCGCCCAACTGTTCAGCCCAATGCTGGCGAGGACAGCTCAGGTG 1091
1095 CATTTGTGTCAACGCTTGAAGATATACAGGAAATATGTTGAGAAAGATGCTGCCCTTTGA 1154
1092 CATCGCGCCACCAACGCTGGAGGAGTACCGCAAGTACGTTGGAGAGGACGCGAGCTTGA 1151
1155 GAGGAGTTTCCAAACAAAGTCTATGTTGCGGAGCCAAAGTGTGCTGACACCATTAGTATCT 1214
1152 GCGGCGGTTCCAGCAGGTGTTCTGTCGCGAGCGGAGCTGCCCGACACCGTCAGCATTTCT 1211
1215 TAGAGACTCAAGGAGAGATATGAGGACATCATGCTGTGCGAAATCCAAAGACAGAGCTCT 1274
1212 GAGGGACTCAAGGAGAGATACGAGGGCACCATGCGGTGAGGATCCAGGACCGCGCCCT 1271
1275 TATAATGTGCTCAGCTGTCTGCTGTACATACTGTCGGCATTTTACCGGATAAAGC 1334
1272 CGTGGTCGCGGCGACAGCTCTCCGCGAGGTACATCATGGTTCGCGACCTGCCCTGACAAAGC 1331
1335 AATTGATTGTTGATGAGGCTTGTGCGAATGTGAGAGTCCAGCTTGTAGTCAACCTGA 1394
1332 CATAGACTGTTGGAGAGGCTTGCCTGCGCAATGTGAGGGTGCAGCTCGACAGCCCGGA 1391
1395 AGAGATTGATACTTTGAAGAGAGAGATGAGCTGGAATTTGAATCTTACGCCCTTGA 1454
1392 GAGATTGATACTCTGAGAGAGAGAAATCCAGCTTGAGGTTGAGCTCCACGCGCTGA 1451
1455 AAGGAGAGAGGATAAAGCCAGCAAGCTCGACTTATAGAGGTGCGGAAAGAGCTTGATGA 1514
1452 GAAGGAGAGAGACAGGCCAGCAAGCCCGCTGATTGAGGTGAGGAGGAAATTTGACGA 1511
1515 CTTGAGAGCAAGCTTTGAGCTCTCAAGTGAATAACAGAAAGAGGAAAGAGAGAAATGA 1574
1512 TCTGAGGACAGCTCGAGCCCTGACCATGAATANNNNNNNNNNNNNNNNNNNNNNNNNN 1571
1575 TGAGATTGAGAGCTTAAACAGAAAGAGAGAGCTCATGTTTCTTTTCAGAGGCGCAGA 1634
1572 TGAGATCAGGAAGCTGAAGCAGCGCGGAGAGCTCCAGTTTCACTTCGAGGAGCGCGA 1631
1635 ACCAAGATATGACTTTGCAAGAGCTGCTGATCTAAGATATGCGCAATTTCAAGAAGTGA 1694
1632 GCGCGATGAGCTGCGCGGTGTGCGCGACTCAAGTACGCGGCGCTCCAGGAATCGA 1691
1695 ATCTGCAATTTGCCCACTTTGAAGGAACTTTCTTGAAGAGAAATGTGATGCTCACAGAAA 1754
1692 GCGCGGATCTCCAAGCT---NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1748
1755 CGTTGGGCTGGAACAATTTGCTGAGGTTGTGAGCGGTTCGACAGGATTTCCAGTGCAGAG 1814
1749 CGTGGCGCTTGAACAAATTTGCAAGAGTGTGAGCGGTTCGAGCGGATTTCCAGTGCACCG 1808
1815 ACTTGGCCAAATAGAGAGAGGTTGATTGCTTCTTGTGATAGGTTGCATAAGCGGCT 1874
1809 GCTTGGCCAGACGACAGAGAGGCTGTGCGCTGCTGACAGGCTTCAACGAGGCT 1868
1875 TGTGGGACAGATCAAGCGGTAAATGACGTTTGTAGGCAATTTTAAGTCAAGGCGAG 1934
1869 GGTGCGCAGACAGAGGCTGTGAGCGCGTTCGAGAGCGGCTGTGAGTTCGAGGCGCG 1928
1935 ACTTGTGAGGACAAACAGCCCAATCGATCATTTTATTTCTTGGACCAACTGTTGTTGG 1994

ULT 2
6643
S
NITIC
SSION

Query Match	19.5%;	Score 605.4;	DB 9;	Length 633;
Best Local Similarity	99.8%;	Prod. No. 8.8e-128;		
Matches 606;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
1	AAAGTATCAATTTTACAACTTACCGCTATTAATCTGTGATTCTCTGC	AAAAGAGAA 60		
27	AAAGTATCAATTTTACAACTTACCGCTATTAATCTGTGATTCTCTGC	AAAAGAGAA 86		
61	GACTTTTTACCGAAGAGAGTCTCTGGCTCATTTGAAGAAA	CTCAACGAACAAACCCGAG 120		
87	GACTTTTTTACCGAAGAGAGTCTCTGGCTCATTTGAAGAAA	CTCAACGAACAAACCCGAG 146		
121	TTCTCATATATCGTTTTTAAGGAGTACAAAAGCTAATCGAAGATGA	ATTCACAGAGAAATTC 180		
147	TTCTCATATATCGTTTTTAAGGAGTACAAAAGCTAATCGAAGATGA	ATTCACAGAGAAATTC 206		
181	ACACAAGACAAACGAGACAATTTGCTACAGCTCATGAGCTAGCTGTGAAT	TCGACGAGACAT 240		
207	ACACAAGACAAACGAGACAATTTGCTACAGCTCATGAGCTAGCTGTGAAT	TCGACGAGACAT 266		
241	GCTCAATTCATCTCTTTTGCATTTAGCTGGTGTTTGAATCTCTGAT	TCCGACCGGTATATTT 300		
267	GCTCAATTCATCTCTTTTGCATTTAGCTGGTGTTTGAATCTCTGAT	TCCGACCGGTATATTT 326		
301	CCTCAAGCAATCTCTAGTGCCTGTCGCGAGAACGACGCTCAATCTGCT	GTGAAGAGTGCATC 360		
327	CCTCAAGCAATCTCTAGTGCCTGTCGCGAGAACGACGCTCAATCTGCT	GTGAAGAGTGCATC 386		
361	AATCAAGCCTTGAAGAGCTTCTTTCAAACTCTCTCCACCTGATGATAT	TTTCAGCGAGT 420		
387	AATCAAGCCTTGAAGAGCTTCTTTCAAACTCTCTCCACCTGATGATAT	TTTCAGCGAGT 446		

RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: meeki@rc.riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified lambda phage vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified paluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

URES Location/Qualifiers
 source
 1. .641
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAF15-30-A07"
 /tissue_type="mixture of silique and flower"
 /lab_host="DH10B"
 /clone_lib="RAF15"
 /note="Site 1: BamHI; Site 2: SalI"
 { COUNT 190 a 146 c 128 g 174 t 3 others
 LIN

ery Match 18.0%; Score 560; DB 9; Length 641;
 est Local Similarity 99.3%; Pred. No. 2.2e-117;
 tches 571; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

1 AAAGTTATCAATTTTCAACATPACCGCTATATCTGTTGATCTCTGCAAAAGAGAA 60
|||||
61 AAAGTTATCAATTTTCAACATPACCGCTATATCTGTTGATCTCTGCAAAAGAGAA 120
|||||
61 GACTTTTACCAGAGAAAGTCTCTGGCTCATTTGAAGAACTCAACGAAACAAACCCAG 120
|||||
121 GACTTTTACCAGAGAAAGTCTCTGGCTCATTTGAAGAACTCAACGAAACAAACCCAG 180
|||||
121 TTCTCATATATGTTTTAAGGGATTAACAAAGCTAATCGAAGATGATCCAGAAATTC 180
|||||
181 TTCTCATATATGTTTTAAGGGATTAACAAAGCTAATCGAAGATGATCCAGAAATTC 240
|||||
181 ACACACAGACAAACGAGCAATTCATGCTAGCTCATGAGCTAGCTGTAATCGAGCAT 240
|||||
241 ACACACAGACAAACGAGCAATTCATGCTAGCTCATGAGCTAGCTGTAATCGAGCAT 300
|||||
241 GCTCAATTCACCTTTTGCATTTAGCTGGTGTGATCTGATCCCAACCGGTATATT 300
|||||
301 GCTCAATTCACCTTTTGCATTTAGCTGGTGTGATCTGATCCCAACCGGTATATT 360
|||||
301 CCTCAAGCAATCTTAGTCCGGTGGCGAGACGAGCTCAATCTGCTGAAGAGTGATC 360
|||||
361 CCTCAAGCAATCTTAGTCCGGTGGCGAGACGAGCTCAATCTGCTGAAGAGTGATC 420
|||||
361 AATCAAGCTTTGAAGAGCTTCTTCCACCAATCTCTCCACCTGATGATATCCAGCGAT 420
|||||
421 AATCAAGCTTTGAAGAGCTTCTTCCACCAATCTCTCCACCTGATGATATCCAGCGAT 480
|||||
421 TCTAGTCTTTAAGGTCATTCGTGCTGCTCAAGCTGCTCAGAAGTCAAGGTTGATCT 480
|||||
481 TCTAGTCTTTAAGGTCATTCGTGCTGCTCAAGCTGCTCAGAAGTCAAGGTTGATCT 540
|||||
481 CATTCGGCTGTTGACCAAGTGTATTAT-GGGTCTTTTGAAGATTCCTCAATCAGGGATT 539
|||||
541 CATTCGGCTGTTGACCAAGTGTATTAT-GGGTCTTTTGAAGATTCCTCAATCAGGGATT 600
|||||
540 GTTGAACGAGTCCGTTGACGACGCGGAGGTAA 574
|||||
601 GTTGAACGAGTCCGTTGACGACGCGGAGGTAA 635

```

LT 5
 4504
 S

CB894504 846 bp mRNA linear EST 24-APR-2003

DEFINITION EST647296 HOGA Medicago truncatula cDNA clone HOGA-31M8, mRNA sequence.
 ACCESSION CB894504
 VERSION CB894504.1 GI:30101673
 KEYWORDS EST.
 SOURCE Medicago truncatula (barrel medic)
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurooids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
 1 (bases 1 to 846)
 Hahn, M.G., Ojanen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S., Utterback, T., Cho, J., and Fraser, C.M.
 ESTs from roots of Medicago truncatula treated with oligogalacturonides of DP 6-20
 Unpublished
 Contact: Michael G. Hahn
 Complex Carbohydrate Research Center
 University of Georgia
 220 Riverbend Road, Athens, GA 30602-4712, USA
 Tel: 706-542-4457
 Fax: 706-542-4412
 Email: hahn@ccrc.uga.edu
 TIGR sequence name: MTMD776TX
 More information is available at: www.medicago.org
 Seq primer: SKmod (CTA GAA CTA gtg gat CC).
 Location/Qualifiers
 1..846
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="A17"
 /db_xref="taxon:3880"
 /clone="HOGA-31M8"
 /tissue_type="3 day old seedling roots"
 /dev_stage="24 hours after treatment in the dark at 26 C with 0.5 mg/ml oligogalacturonides (DP 6-20) in the presence of 100 ug/ml Gentamicin"
 /lab_host="XLOLR"
 /clone_lib="HOGA"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist helper phage and propagated in SOLR cells."
 BASE COUNT 270 a 139 c 243 g 194 t
 ORIGIN
 Query Match 17.5%; Score 543.4; DB 14; Length 846;
 Best Local Similarity 78.7%; Pred. No. 1.6e-113;
 Matches 665; Conservative 0; Mismatches 171; Indels 9; Gaps 1;
 QY 1327 GATTAACCAATTGATTTGGTTGATGAGCTGTGCGAATGTGAGAGTCCAGTTGATGT 1386
 DB 1 GACAAGGCCATTGATTTAGTTGATGAGGCGCATGTGCTTAATGTGAGAGTTCAACTTGATGT 60
 QY 1387 CAACCTGAAGAGATTGATAACCTTGAAGCAAGAGGATGCGCTGGAAATTTGAACCTTCAC 1446
 DB 61 CAACCTGAAGAAATTGCAATCTTTGAAGAGAGAGATTCGCAATTAGAGTTGAACCTGCAT 120
 QY 1447 GCCTTGGAAAGGGAGAGAGATTAAGCCAGCAAGCTCGACTTATAGAGGTGCGGAAGAG 1506
 DB 121 GCTCTAGAGAAAGAGAGAGGACAAAGCTAGCAAGCCCGCTCTTGTAGATGTACCGAGAGAA 180
 QY 1507 CTTGATGACCTGAGAGCAAGCTTCAGCCTCTCAGATGAAATACAGAAAGGAGAGAG 1566
 DB 181 CTTGATGACTTGAAGAGCAAGCTTCACCTCTCGAAGTGAAGTATAGTAAAGAGAGAG 240
 QY 1567 AGAATTCATGAGATTTCGAAGGCTTAAACAGAAAGAGAGAGAGCTCATGTTTCTTTTCGAG 1626
 DB 241 AGGATTGATGAGATTTCGAAGGCTTCAGCAGAAAGAGAGAGCTCTCTTTTTCGATACAG 300

1627 GAGCGAGAAAGAGATATGACCTTGGCAAGAGCTGCTGATCTTAAGATATGGCGCAATTCAA 1686
 301 GAGCGAGAGAGCGGTATGATCTAGCAGAGAGCTGCGACCTGCGATATGGTGCATTTGAA 360
 1687 GAAGTGGAACTGCAATTGGCCCAACTTTGAAGAACTTCT-----TCTGAAGAGAAAT 1737
 361 GAGGTGGAACTGCAATTAAAAATCTTTGAAGGTAGCACTGATGGGAAACACGGATGAGAAC 420
 1738 GTGATGCTCAGAGAAAGCTTGGCGCTGAAACACATCTGCTGAGGTGTGAGCCGTGGACA 1797
 421 TTAATGTTGACGGAACAGATTGGACCGGACCAATAGCCGAGGTGTGAGCCGATGGAAT 480
 1798 GGGATTCAGTGCAGAGACTTGGCCAAATGAGAGAGAGAGGTGATTTGCTTCTGCTGAT 1857
 481 GGTATACAGTACTAGCTCGGCCAAATGAGAAAGCAAGGTGTGTTGACTTGGTGAC 540
 1858 AGGTGCTATAGCGGTGTGGGACAGAACTCAAGCGGTAATGACAGTCTTCTGAGGCAATT 1917
 541 AGACTGCACACTAGAGTGTGGGACAGACCAAGCGGTTAATGCTGTGTTGCTGAGGCTGTA 600
 1918 CTAAGTCAAGGCGAGGACTTGTAGGGCAACACAGCCCACTGGATCATTTCTTATTCCTT 1977
 601 TTGAGATCAAGAGCTGTTTGGTAAAGTCTGAGCTTCAAGGCTCTTTCATTTATTCCTT 660
 1978 GGACCAACTGTTGTCGAACTGAGCTGCCAAGGCTCTTGTCTGAGCAGCTGTTTGTAT 2037
 661 GGTCCGACTGTTGTTGTTAAAGTCTGAGCTTCAAGGCTCTTTCAGAGCAGCTATTGAT 720
 2038 GATGAACACTCTTAGTTCGGAATGATGATGTCGGAATATATGGAACAACTCTGTCTCT 2097
 721 GATGAACACTCTTAGTTCGGAATGATGATGTCGGAATATATGGAACAACTCTGTATCA 780
 2098 CGCTCTATTTGGGCAACCAAGGCTATGCTTCCACGAGGAAGTGTGACAACTAACTGAG 2157
 781 AGATGATGTTGTCACCCAGGCTATGTTGGACATGAGGAAGGGGTCATTTAACAGAA 840
 2158 GCTGT 2162
 841 GCTGT 845

ULT 6
 50711
 US
 INITIATION
 BOGAQ17TF BOGA Brassica oleracea genomic clone BOGAQ17, genomic
 survey sequence.
 BH250711
 BH250711.1 GI:17074863
 GSS.
 Brassica oleracea
 ORGANISM
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; euroside II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 818)
 Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.
 Location/Qualifiers
 1. 818
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"

TURES
 source

/db_xref="taxon:3712"
 /clone="BOGAQ17"
 /clone_lib="BOGA"
 /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pHOS1 using BstXI linkers"
 BASE COUNT 222 a 158 c 217 g 221 t
 ORIGIN
 Query Match 17.5%; Score 542; DB 28; Length 818;
 Best Local Similarity 85.7%; Pred. No. 3.2e-113;
 Matches 615; Conservative 0; Mismatches 100; Indels 3; Gaps 1;
 1439 AACTTCACGCCCTTGGAAAGGAGAGATTAAGCAGCAAGAGCTCGACTTATAGAGTGC 1498
 77 AACTTCTCTTTTGGGCAAGTCTCTGATATTGTTATCTTTGAATTAATGATCTGTAGGTGC 136
 1499 GGAAGAGCTTATGACCTGAGAGACAAGCTTCAAGCTTCAAGATGAATACAGAAAGG 1558
 137 GGAAGAGCTTATGACCTGAGGGAAGCTTCAAGCTTCAAGATGAATACAGAAAGG 196
 1559 AGAAGAGAGAAATTCATGAGATTTCGAAGCTTTAAACAGAAAGAGAGCTCATGTTT 1618
 197 AGAAGAGAGAAATTCATGAGATTTCGAAGCTTTAAACAGAAAGAGAGCTCATGTTT 256
 1619 CTTTCAGGAGCAGAACGATATGACCTTTCAGAGCTGCTGATCTAAAGATATGCG 1678
 257 CTTTCAGGAGCAGAACGATATGACCTTTCAGAGCTGCTGATCTAAAGATATGCG 316
 1679 CAATTCAAGAGTGAATCTGCAATTCGCAACTTGAAGGACTTCTTCTGAAGAGATG 1738
 317 CAATTCAGAGTGAATCTGCAATTCGCAACTTGAAGGACTTCTTCTGAAGAGATG 373
 1739 TGATGCTCAGCAAAAACGTTGGGCTTGAACACATTTGCTGAGCTTGTGAGCCGTTGGA 1798
 374 TGATGCTCAGCAAAAACGTTGGGCTTGAACACATTTGCTGAGCTTGTGAGCCGTTGGA 433
 1799 GGAATTCAGTGCAGAGACTTGGCCAAATGAGAGAGAGGTTGATGTTGCTGATG 1858
 434 GGAATTCAGTGCAGAGACTTGGCCAAATGAGAGAGAGGTTGATGTTGCTGATG 493
 1859 GGTTCATAGCGGCTTGTGGACAGATCAAGCGTAAATGCGAGTTTCTGAGGCAATTC 1918
 494 GGTTCATAGCGGCTTGTGGACAGATCAAGCGTAAATGCGAGTTTCTGAGGCAATTC 553
 1919 TAAGTCAAGGCGCAGACTTGTGAGGCAACACAGCCAACTGGATCATTTTATTCCTG 1978
 554 TAAGTCAAGGCGCAGACTTGTGAGGCAACACAGCCAACTGGATCATTTTATTCCTG 613
 1979 GACCAACTGCTGTTGGCAAACTGAGCTGCGCAAGCTTCTGAGCAGAGCTGTTGATG 2038
 614 GACCAACTGCTGTTGGCAAACTGAGCTTCTGAGCAGAGCTTCTGAGCAGAGCTTCTG 673
 2039 ATGAAACCTCTTAGTTCGATTTGATGATGTCGGAATATATGGAACAACTCTGCTCTC 2098
 674 ATGAAACCTCTTAGTTCGATTTGATGATGTCGGAATATATGGAACAACTCTGCTCTC 733
 2099 GCCTCATTTGGGCAACCAAGGCTGTTGTTGTCAGAGAAAGGTCGACAACTAACTGA 2156
 734 GCCTCATTTGGGCAACCAAGGCTGTTGTTGTCAGAGAAAGGTCGACAACTAACTGA 791

RESULT 7
 BH465756/c
 LOCUS
 DEFINITION
 BOGH104TF BOGE Brassica oleracea genomic clone BOGE104, genomic
 survey sequence.
 BH465756
 BH465756.1 GI:17664770
 GSS.
 Brassica oleracea
 ORGANISM
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eucosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 863)
 Town, C.D., Van Aken, S., Uterback, T., Koo, H. and Fraser, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished
 Other_GSSs: BOGEI04TR
 Contact: Chris Town

TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

URES
 source
 Location/Qualifiers

1..863
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="T0100DH3"
 /db_xref="taxon:3712"
 /clone="BOGEI04"
 /clone_lib="BOGE"

/note="Vector: pHOS1; Site 1: BatXI; 2-3 kb sheared
 genomic DNA inserted into pHOS1 using BstXI linkers"

COUNT 213 a 234 c 165 g 251 t

Query Match 17.4%; Score 541.4; DB 28; Length 863;
 Best Local Similarity 87.2%; Pred. No. 4.5e-113;
 Matches 606; Conservative 0; Mismatches 86; Indels 3; Gaps 1;

1439 AACTTCACGCCCTTGAAGGGAGAGGATAAGCCAGCAAGCTCGACTTATAGAGTGC 1498

697 AACTTCCTTTGGGCAAGTCTTGATTTGTTATCTGATATGATCTGAGTGC 638

1499 GGAAGAGCTTGATCACTGAGAGCAAGCTTCAGCCTTCACGATGAATACAGAAAGG 1558

637 GGAAGAGCTTGATCACTGAGGAGCAAGCTTCACGATGAATACAGAAAGG 578

1559 AGAAGAGAGATTGATGATTCGAAGGCTTAACAGAAAGAGAGAGCTCATGTTT 1618

577 AGAAGAGAGATTGATGATTCGACGCTTAAACAGAAAGAGAGAGCTCATGTTT 518

1619 CTTTCAGAGAGCAGAAACGATATGATCTGCAAGAGCTGCTGATCTAAGATATGCG 1678

517 CTTTCAGAGAGCAGAAACGATATGATCTGCAAGAGCTGCTGATCTAAGATACGAG 458

1679 CAATTCAAGAGAGTGAATCTGCAATTGCCCACTTGAAGGAACTTCTTCTGAAGAGATG 1738

457 CAATCCAAGAGTGAATCTGCAATAGCCCACTCGAAGGAA---CTTCTGACGAGACA 401

1739 TGATGCTCAGAGAAACCTTGGGCTTGGAACATGCTGAGGTTGTGAGCGGTGGACAG 1798

400 TGATGCTCAGAGAAACCTTGGGCTTGGAACATGCTGAGGTTGTGAGCGGTGGACAG 341

1799 GGATTCAGAGTGAATCTGCAATTGCCCACTTGAAGGAACTTCTTCTGAAGAGATG 1858

340 GGATTCAGAGTGAATCTGCAATTGCCCACTTGAAGGAACTTCTTCTGAAGAGATG 281

1859 GGTTCATTAAGCGGTTTGGGACAGATCAAGCGGTAATGCGATTTCTGAGGCAATTC 1918

280 GGTTCATTAAGCGGTTTGGGACAGATCAAGCGGTAATGCGATTTCTGAGGCGATCT 221

1919 TAAGTTCAGGCGAGGCTTGTAGGCGACACACCGCAACTGGATCATCTTATCTCTTG 1978

220 TAAGTTCAGGCGAGGCTTGTAGGCGACACACCGCAACTGGATCATCTTATCTCTTG 161

1979 GACCAACTGGTTCGGCAAAACTGAGCTCGCAAGGCTTCTTCTGAGCAGCTGTTTGATG 2038

160 GCCCAACTGGTTCGGCAAAACTGAGCTTCTTCTGAGCAGCTTCTTCTGATG 101

2039 ATGAAACCTCTTAGTTCGGATTGATGTCGGAAATATATGGAACAACACTCTCTCTCTC 2098

Db 100 ATGAAACCTCTTAGTTCGGATTGATGTCGGAAATATATGGAACAACACTCTCTCTCTC 41
 Qy 2099 GCCTCATTTGGGCAACCAACAGGGTATGTTGGTCAAC 2133
 Db 40 GCCTCATTTGGGCAACCAACCTGGGTAAGTACCTAAC 6

RESULT 8
 LOCUS CB894478 837 bp mRNA linear EST 24-APR-2003
 DEFINITION EST647270 HOGA Medicago truncatula cDNA clone HOGA-31G20, mRNA
 sequence.
 ACCESSION CB894478
 VERSION CB894478.1 GI:30101647
 KEYWORDS EST.
 SOURCE Medicago truncatula (barrel medic)
 ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eucosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 1 (bases 1 to 837)
 Hahn, M.G., Ojanen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S.,
 Uterback, T., Cho, J. and Fraser, C.M.
 ESTs from roots of Medicago truncatula treated with
 oligogalacturonides of DP 6-20
 Unpublished

REFERENCE
 AUTHORS Contact: Michael G. Hahn
 Complex Carbohydrate Research Center
 University of Georgia
 220 Riverbend Road, Athens, GA 30602-4712, USA
 Tel: 706-542-4457
 Fax: 706-542-4412
 Email: hahn@ccrc.uga.edu

TIGR sequence name: MTMD746TK
 More information is available at: www.medicago.org

Seq primer: SKmod (CTA GAA CTA gtg gat CC).

FEATURES
 Location/Qualifiers

1..837
 /organism="Medicago truncatula"

/mol_type="mRNA"

/cultivar="A17"

/db_xref="taxon:3880"

/clone="HOGA-31G20"

/tissue_type="3 day old seedling roots"

/dev_stage="24 hours after treatment in the dark at 26 C

presence of 0.5 mg/ml oligogalacturonides (DP 6-20) in the

presence of 100 ug/ml Gentamicin"

/lab_host="XLOLR"

/clone_lib="HOGA"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
 was directionally ligated into the Unizap XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in SOLR cells."

BASE COUNT 258 a 147 c 236 g 196 t

ORIGIN

Query Match 17.0%; Score 529; DB 14; Length 837;
 Best Local Similarity 78.0%; Pred. No. 3.1e-110;
 Matches 653; Conservative 0; Mismatches 175; Indels 9; Gaps 1;

Qy 1496 TCGGGAAGAGCTTGATGACCTGAGAGACAAGCTTCAGCTTCACGATGAATACAGAA 1555

Db 1 TACGGAGAGAACTTGATGACTTGGAGAGACAAGCTTCAACCTCTGAGATGAAGTAGTA 60

Qy 1556 AGGAAAGAGAGAAATGATGATTCGAGGCTTAACAGAAAAGAGAGAGCTCATGT 1615

Db 61 AAGAGAAAGAGAGAGATGATGATTCGAGGCTTAAGCAGAAACGTAAGAGAGCTCTCT 120

Db 539 GATGAAGTAGAGAAAGCACATATCGCGGTATTCAATACCTCCTCCAGGTTTAGACGAT 598

Tel: 530 754 6561

UC Davis, Plant Pathology
One Shields Ave. Davis, CA

2254 GGTGATTGACAGAGGCGCAGGACAGTCTGATTCAGGACTCGGTGATTAATCATG 2313
 599 GGAAGGTTAACTGATGATGCGCAGGCGGACTGTAGACTTCAAAACACAGTGATCATATG 658
 2314 ACATCAAACTTGGTGTCTGACACCTCTCTGACGGGCTAACTGGGAAAGTAAACAATGGAA 2373
 659 ACTTCAAACTTGGAGCTGAGCATCTCTCTCGGGTTTAGTAGGCAAGTGATCAATGCAA 718
 2374 GTGCGCCGGGACTGTGTGATCGGGAGGTGAGGAAACACTTCAGACAGAGCTCTTGAAC 2433
 719 GATGCTCGTGACCGGGTGATGACAGAGGTGAGNAGGCACTTCAGGCCAGAGTTGCTTTAAC 778
 2434 AGGCTTACGACGAGATTGTGCTGCTCGACCCCTTTACATGACCAAGTTGAGGAAAGTAGCT 2493
 779 CGGTGATGAAATGTGTTGATTTGATTCCTCTCATCATGATCACTGAGAAAGTTGCC 838
 2494 CGGCT 2498
 839 AGGCT 843

JLT 10

48385

US

INITIATION

SSION

ION

WORDS

ANCE

THORS

TITLE

JOURNAL

MENT

URES

source

BG648385 736 bp mRNA linear EST 24-APR-2001
 EST510004 HOGA Medicago truncatula cDNA clone phOGA-19L1 5' end,
 mRNA sequence.

BG648385

BG648385.1

GI:13783497

EST.

Medicago truncatula (barrel medic)

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

1 (bases 1 to 736)

Hahn, M.G., Ojanen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S.,

Utterback, T., Cho, J., and Fraser, C.M.

ESTs from roots of Medicago truncatula treated with

oligogalacturonides of DP 6-20

Unpublished

Contact: Michael G. Hahn

Complex Carbohydrate Research Center

University of Georgia

220 Riverbend Road, Athens, GA 30602-4712, USA

Tel: 706-542-4457

Fax: 706-542-4412

Email: hahn@ccrc.uga.edu

G391991e TIGR sequence name: MTMCK61TK More information is

available at: www.medicago.org

Seq primer: SKmod (CTA GAA CTA gtg gat CC).

Location/Qualifiers

1. 736

/organism="Medicago truncatula"

/mol_type="mRNA"

/cultivar="A17"

/db_xref="taxon:3880"

/clone="phOGA-19L1"

/tissue_type="3 day old seedling roots"

/dev_stage="24 hours after treatment in the dark at 26 C

with 0.5 mg/ml oligogalacturonides (DP 6-20) in the

presence of 100 ug/ml Gentamicin"

/lab_host="XLOLR"

/clone_lib="HOGA"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:

XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA

was directionally ligated into the Unizap XR vector from

Stratagene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised

from the recombinant lambda-zap phage using Ex-assist

helper phage and propagated in SOLR cells."

201 a

88 c

219 g

228 t

COUNT

ORIGIN

Query Match 16.3%; Score 505.8; DB 10; Length 736;
 Best Local Similarity 81.3%; Pred. No. 6.2e-105;
 Matches 599; Conservative 0; Mismatches 137; Indels 1; Gaps 1;

QY 451 CAAGCTGCTCAGAGTTCAGAGGTGATCTCACTATTGGCTGTGACCACTGATTAATGGGT 510
 Db 1 CAAGCTGCTCAGAA-TTACGTGTGATCTCACTATTAGCTGTGATCACTTAATTTGGGA 59
 QY 511 CTCTTTGAAGATTCTCAATCAGGGATTGTTGAAACGAGTCCGTGTAGCCAGCGCGAGG 570
 Db 60 ATTCTTTGAGATTCTCAGATTGCTGATTGTTTAAAGAGCTGTGTTGCTTTCTAGG 119
 QY 571 GTAAAGTCTGAGGTTGAGAAAGCTTCGTGGGAAAGGAAAGGAAAGTTGAGAGTCTTCA 630
 Db 120 GTTAAAGCTGAGGTTTGAAGAGCTGAGGGGTAAAGATGGGAAAGGTTTGAAGTCTTCC 179
 QY 631 GGGGACACAAATTTTCAAGCTTTAAAGACTTATGCAAGAGATTGTTGTTGAGCAACAGG 690
 Db 180 GGTGATACGAATTTTCAAGCATTTGAAGACTTATGTTAGAGATTGTTGTTGAAACAGAG 239
 QY 691 AAGCTTGATCTGTGATTTGGTCTGATGAGGAGATTAGAAGAGTCTGAGGATCTTTTCG 750
 Db 240 AAGCTTGATCTGTGTTATTGGACCGCAGCAGAGAGATTGCAAGAGTTGTCAGGATTTTCA 299
 QY 751 AGGAGACGAGAAACAATCTCTGTGCTTATTGAGAGCCAGAGTTGGTAAACAGAGTGTG 810
 Db 300 AGGAGGACTTAAGATAATCCAGTCTTATTGTTGAGCTGTGTTGGGAAAGTCTGCTTT 359
 QY 811 GTTGAAGGTTTGAACAAAGGATTCTGAAAGGAGATGTCGCCAAGTCTTACTGATGTG 870
 Db 360 GTTGAAGGTTGGCTCAGAGATTGTTAGAGTGATGTTCTGCAATCTTGTGCTGATGTT 419
 QY 871 AGATTAAATTTGTTGGACATGGGTGCTGTAGTGTGCTGTGCTGCTGCTGCTGCTGCTG 930
 Db 420 AGTTAAATTTGCTCTGATATGGGAGCATTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479
 QY 931 GAAGAAAGTTGAAATCTGTTTGAAGAGTTGAGGAGCTGAGGCAAGTGAATGATTTCTC 990
 Db 480 GAAGAGAGTTGAAGGCTGTTTGAAGAGTTGAAGAGCTGAAGGAGGTTAATTTCTT 539
 QY 991 TTATTGATGAGATTCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1050
 Db 540 TTCATGATGAGATTCACTTGTCTTGGAGCTGTGAGACAGAGTTCGATGATTTGGTGTAC 599
 QY 1051 GCTAATCTTTCAAGCCCATGTTAGCTAGAGGGCAGCTTCGATGCTGCTGCTGCTGCTGCTG 1110
 Db 600 GCTAATCTTTTAAAGCAATGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659
 QY 1111 CTGGAAGATACAGGAAATATGTTGAGAAAGATGCTGCTTGTGAGAGAGGTTCCAAAC 1170
 Db 660 CTGGAAGATATAGGAAGTATGTTGAAAGAGATGCGCATTTGAAAGAGGTTTCAACAG 719
 QY 1171 GTCTATGTTGGGAGGCC 1187
 Db 720 GTTATGTTGGTGTGACC 736

RESULT 11

CB893989

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Medicago truncatula (barrel medic)

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

CB893989 812 bp mRNA linear EST 24-APR-2003
 EST5646781 HOGA Medicago truncatula cDNA clone HOGA-29N13, mRNA
 sequence.

CB893989

CB893989.1

GI:30101158

EST.

Medicago truncatula (barrel medic)

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 812)
Hahn,M.G., Ojanen-Reuhs,T., Samac,D., Town,C.D., Van Aken,S.,
Utterback,T., Cho,J. and Fraser,C.M.
ESTs from roots of Medicago truncatula treated with
oligogalacturonides of DP 6-20
Unpublished
Contact: Michael G. Hahn
Complex Carbohydrate Research Center
University of Georgia
220 Riverbend Road, Athens, GA 30602-4712, USA
Tel: 706-542-4457
Fax: 706-542-4412
Email: hahn@ccrc.uga.edu
TIGR sequence name: MTMDM797K
More information is available at: www.medicago.org
Seq primer: SKmod (CTA GAA CTA gta gat CC).
Location/Qualifiers
1..812

FEATURES

Source
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="HOGA-29N13"
/tissue_type="3 day old seedling roots"
/dev_stage="24 hours after treatment in the dark at 26 C
with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
presence of 100 ug/ml Gentamicin"
/lab_host="XLOLR"
/clone_lib="HOGA"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI. cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in SOLR cells."
258 a 137 c 233 g 184 t

BASE COUNT
GIN

Query Match 16.2%; Score 502.4; DB 14; Length 812;
Best Local Similarity 77.2%; Pred. No. 3.8e-104;
Matches 627; Conservative 0; Mismatches 176; Indels 9; Gaps 1;
1420 AGGATGACGCTGGAATTAATCACTTACGCTTGGAAAGGAGAGATTAAGCCAGAAA 1479
1 AGAATGCAATTAGAAAGTTGAATGCTGCTAGAGAAAGAGAGAGAGAGAGAGAGAG 60
1480 GCTGACCTATAGAGCTGCGAAAGAGCTTGTGACCTGAGAGCAAGCTTCAGCCTC 1539
61 GCCCTCTGTAGATGTACGGAGAGAACTTGAAGCTTGTGAGAGCAAGCTTCAACCTCTG 120
1540 ACGATGAATACAGAAAGAGAGAGAGAGAGATTTGATGAGATTGGAAGCTTAAACAGAAA 1599
121 AAGATGAAGTATAGTAAAGAGAGAGAGAGAGATTGATGAGATTGGAAGCTGAAGAGAA 180
1600 AGAAGAGAGCTCATGCTTTTCTTTGAGAGAGAGAGAGAGAGATGATGAGAGCT 1659
181 CTTGAGAGAGCTCTCTTTGCACTACAGAGAGAGAGAGAGAGCGGTATGATCTAGCAGAGCT 240
1660 GCTGATCTAAGATATGGCGCAATTCAGAGAGTGAATCTGCAATTTGCCCACTTGAAGGA 1719
241 GCGGACCTCGATATGGTGAATTCAGAGAGTGGAACTGCAATTTAAATCTTGAAGGT 300
1720 ACTTCT-----CTGGAAGAGATGTGATGCTTCACAGAAAAGTTGGCGCTTGAACAC 1770
301 AGCACTGATGGGAACACGATGAGAACTTAATGTTGACGGAACAGTTGGACCGAGCA 360
1771 ATTGCTGAGGTGTGAGCGGTGGAGAGAGATTCAGTGCAGAGACTTGGCCAAATGAG 1830
361 ATAGCCGAGGTGTTAGCCGATGGAGCTGGTATACCAAGTGCATAGCTCGGCCAAATGAG 420
1831 AAGGAGAGGTGATTTGGTCTTCTGCTATAGGTTGCATAGCGGGTTGTGGGACAGATCAA 1890

Db 421 AAGCAAAGTTGGTTGGACTTGGTGACAGACTGCACACTAGAGTCTGTGGGACAGACCAA 480
QY 1991 GCGTAATATCAGATTTCTGAGGCAATCTAAGGTCAAGGCGAGAGACTTGTAGGCGACAA 1950
Db 481 GCGTTAATGCTGTGCTGAGGCTGTATTGAGATCAAGAGCTGTGTTTAGAAGACCCCAA 540
QY 1951 CAGCCAACTGGATCATTTCTTATTCCTTGGACCAACTGGTGTGGCAAACTGAGCTCGCC 2010
Db 541 CAACCAACTGGTTCCTTCTTATTCCTTGGTCCGACTGTGTTGTTAACTGAGCTTGA 600
QY 2011 AAGGCTCTTCTGAGCAGCTGTTTGATGATGAATAAAGCTTCTTGTGCGATTTGATGTCG 2070
Db 601 AAGGCTCTTCTGAGCAGCTGTTTGATGATGAATAAAGCTTCTTGTGCGATTTGATGTCG 660
QY 2071 GAATATATGAACAACACTCTGCTCTGCTCTATTTGGGCGACACACAGGATTTGTTGT 2130
Db 661 GAATACATGAACAACACTCTGCTCTATCAAGATTAATTTGGTGCACACCGGGTATTTGGA 720
QY 2131 CAGGAGAGGTGGACAACTTAATCTGAGGCTGTGAGGAGCGACTTATTGTGTCACTC 2190
Db 721 CATGAGGAAGGGGTCAATTAACAGAGCTGTAAAGGAGAGGCCATACAGTGTGTACTC 780
QY 2191 TTTGATGAATGGAGAGAGCTCATGTTGCTGT 2222
Db 781 TTTGATGAATGGAGAGAGCTCATGTTGCTGT 812
RESULT 12
BH603944/c
LOCUS
DEFINITION
BH603944
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
BH603944 750 bp DNA linear GSS 15-DEC-2001
BOGTL30TF BOGT Brassica oleracea genomic clone BOGTL30, genomic
survey sequence.
BH603944
GSS.
GI:17856390
Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosid II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 750)
Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other GSSs: BOGTL30TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TP
Class: sheared ends.
Location/Qualifiers
1..750
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000H3"
/db_xref="taxon:3712"
/clone="BOGTL30"
/clone_lib="BOGT"
/note="Vector: pHO1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHO1 using BstXI linkers"
BASE COUNT 154 a 236 c 150 g 210 t
ORIGIN

Query Match 15.9%; Score 494; DB 28; Length 750;
Best Local Similarity 81.3%; Pred. No. 3.2e-102;
Matches 597; Conservative 0; Mismatches 135; Indels 2; Gaps 2;
QY 419 GTTCTAGTCTTATTAGGTCATCTGCTGCTCAAGCTGCTCAGAGTCAAGGTGATA 478

750 GCTCAGAGCTCTATCAAGCGATACGCGCGCTCAAGCAGCTCAGAGTCCGCTGGCGACA 691
 479 CTCATTGGCT-GTTCACAGATTGATTATGGGTCTCTTGAAGATCTCAATCAGGGAT 537
 690 CACACTTAGCCGTGACAGCTGTCTAGAGCCCTCTCGAGACTCTCAGATCAAGAT 631
 538 TTCTTGAACGAAGTGGTGTAGGACGCGGAGGTAACTCTGAGGTGAGAGTTCGT 597
 630 CTTTAAAGAAAGTGGGTGCGAGCGGAGGTGAATCCGAGTTGAGAAGTTGGT 571
 598 GCGAAGAAAGGGAAGAACTTGAGAGTCTTCAAGGGGACACAAATTTCAAGCTTAAAG 657
 570 GGGAGCAGAGGAAGAAAGTGTAGAGCGCTTCGGGGGATACAACTTTCAAGCGTTGAAG 511
 658 ACTTATGGAAGAGATTGGTTGAGCAAGCAGGAGCTTTGATCTCTGTGATTTGGTCTGAT 717
 510 ACTTACGGAAGAGACTTGTGGAAGAGCGGGAAGCTTTGATCTCTGTGATTTGGTCTGAT 451
 718 GAGGAGATTAGAGAGTCTGAGGATCTTTGAGAGAGCAAGAACTCTCTGTGCTT 777
 450 GAGGAGATTAGAGAGTCTGAGGATCTTTGAGAGAGCAAGAACTCTCTGTGCTT 391
 778 ATTGAGAGCCAGGAGTTGCTTAAACAGCTGTGTGAGGTTTGAAGCTTGAAGGATTTG 837
 390 ATCGGAGAGCCAGGTGTGCGGAAACAGCTGTGTGAGGCTTGAAGCTTGAAGGATTTG 331
 838 AAGAGAGTGTCCCAACAGCTTCTTCTGATGAGATTAATTTCTGAGCAATCTCTGTGCT 897
 330 AAGAGAGAGCTCCCAACAGCTTCTTCTGATGAGATTAATTTCTGAGCAATCTCTGTGCT 271
 898 TTAGTTGCTGTGCTTAAATACCGAGAGAGTTTGAAGAGGTTGAAGATCTGTTTGAAG 957
 270 TTGGTGTGCTGTGCTTAAATACCGAGAGAGTTTGAAGAGGTTGAAGATCTGTTTGAAG 211
 958 GAAGTTGAGGAGCTTGAAGGCAAGTGTCTTCTTATTTGATGAGATTCATTTGCTTCT 1017
 210 GAAGTCGAGGAGCTGAGGGAAGTGTCTTCTTATTTGATGAGATTCATTTGCTTCT 151
 1018 GTGTGCGCAAACTGAAGGTCGATGAGTCAAGCTTGAAGATTCATTTGCTTCT 1077
 150 GAGAGTGGAGAGCCGAGGTCGATGAGCTGCGAATCTGTTTGAAGCTTGTGCT 91
 1078 AGAGGAGAGCTTCAAGTGTGCTTGAAGCTTGAAGATTCATTTGCTTCT 1137
 90 AGAGTCAAGTGTGCTTGAAGCTTGAAGATTCATTTGCTTCT 1137
 1138 AAGATGCTGCTT 1151
 31 AAGAGCTGCTT 18

ULT 13
 93376
 US
 INITIATION
 CB893376 783 bp mRNA linear EST 24-APR-2003
 EST64168 HOGA Medicago truncatula cDNA clone HOGA-27L21, mRNA
 sequence.
 CB893376
 CB893376.1 GI:30100544
 EST.
 MEDICATION
 Medicago truncatula (barrel medic)
 Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosoids 1; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 1 (bases 1 to 783)
 Hahn, M.G., Ojanen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S.,
 Utterback, T., Cho, J. and Fraser, C.M.
 ESTs from roots of Medicago truncatula treated with
 oligogalacturonides of DP 6-20
 Unpublished
 Contact: Michael G. Hahn
 Complex Carbohydrate Research Center

University of Georgia
 220 Riverbend Road, Athens, GA 30602-4712, USA
 Tel: 706-542-4457
 Fax: 706-542-4412
 Email: hahn@ccr.uga.edu
 TIGR sequence name: MTWDE71TK
 More information is available at: www.medicago.org
 Seq primer: SKmod (CTA GAA CTA GTG GAT CC).
 Location/Qualifiers

FEATURES
 source

1..783
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="Al7"
 /db_xref="taxon:3880"
 /clone="HOGA-27L21"
 /tissue_type="3 day old seedling roots"
 /dev_stage="24 hours after treatment in the dark at 26 C
 with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
 presence of 100 ug/ml Gentamicin"
 /lab_host="XLOLR"
 /clone_lib="HOGA"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
 was directionally ligated into the Unizap XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in SOLR cells."
 BASE COUNT 219 a 125 c 211 g 228 t
 ORIGIN

Query Match 15.8%; Score 490.2; DB 14; Length 783;
 Best Local Similarity 76.6%; Pred. No. 2.4e-101;
 Matches 600; Conservative 0; Mismatches 183; Indels 0; Gaps 0;
 Qy 193 AACGAGCAATTCCTCAGCTCATGAGCTAGCTGTGAATCGAGGACATGCTCAATTCCT 252
 Db 1 AACGAGCAATTCCTCAGCTCATGAGCTAGCTGTGAATCGAGGACATGCTCAATTCCT 60
 Qy 253 CCCTTGCATTTAGCTGGTGTGATCTCTGATCCACCGGTATATTTCTCAAGCAATC 312
 Db 61 CCCTTGCATTTAGCTGGTGTGATCTCTGATCCACCGGTATATTTCTCAAGCAATC 120
 Qy 313 TCTAGTGCCTGGTGGGAGAACGAGCTCAATCTGCTGAAGAGTGATCAATCAAGCTTGT 372
 Db 121 TCAATGTTGCTGGTGAAGATCAGACAGCTGCGTGAACGTTGTTGAACAGATTA 180
 Qy 373 AAGAAGTCTCTTCAATCTCTCTCCAGTGAATATTCAGCGAGTTCTAGTCTTAT 432
 Db 181 AAGAAGTCTCTTCAATCTCTCTCCAGTGAATATTCAGCGAGTTCTAGTCTTAT 240
 Qy 433 AAGGTCAATTCCTGGTCTCAAGCTCTCAGAGTCAAGAGTCAAGTCAATTCCTTGT 492
 Db 241 AAGCAATTCCTGGTGAAGCAAGCTCTCAGAGTCAAGAGTCAAGTCAATTCCTTGT 300
 Qy 493 GACCACTGATTTAGTGGTCTCTTGAAGATTCCTCAATCAGGATTTGTTGAACGAATC 552
 Db 301 GATCACTTAACTCTGGGATTTCTGAGGATTTCTAGATTTGTTGAAGAGCT 360
 Qy 553 GGTGTAGCAACGCGGAGGTAAAGTCTGAGTTGAGAGCTTCTGTTGAAGAGAGGGAAG 612
 Db 361 GGTGTGCTGTTCTTAGGGTTAAACTGAGGTTGAGAGCTGAGGGTAAAGATGGGAAA 420
 Qy 613 AAAGTTGAGAGTCTTCAAGGAGCAATTTCAAGCTTTAAAGACTTATGAGAGAT 672
 Db 421 AAGTTGAAGTCTTCCGGTGAAGATTTCAAGCAATTTGAAGACTTATGAGAGAT 480
 Qy 673 TTGGTTGAGCAAGCAGGAGGAGCTTCTGATCTGTTGATTTGTTGAGGAGATTAAGA 732
 Db 481 TTGGTTGAGCAAGCAGGAGGAGCTTCTGATCTGTTGATTTGAGGAGGAGATTTCAAGA 540
 Qy 733 GTCTGAGAGATTTCTTTCGAGGAGAAAGCAAGCAATCTGTTCTTATTTGAGAGCAGGA 792

```

e
seq primer: smou (LTA 9AA CIA 9TG GAR CC).
location/Qualifiers
1. ..818
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="PHOGA-23B24"
/tissue_type="3 day old seedling roots"
/dev_status="24 hours after treatment in the dark at 26 C
with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
presence of 100 ug/ml Gentamicin"
/lab_host="XL0LR"
/clone_lib="HOGA"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in SOLR cells."
239 a. 151 C 230 G 198 C
15.5%; Score 483; DB 10; Length 818;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

nucleic - nucleic search, using sw model

on: February 12, 2004, 17:32:36 ; Search time 207 Seconds
(without alignments)
6620.747 Million cell updates/sec

e: US-09-812-350-30

ect score: 3105
ence: 1 aaggtatcaatttacaac.....tctttggggtttttcata 3105

ing table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

ched: 569978 seqs, 220691566 residues

l number of hits satisfying chosen parameters: 1139956

um DB seq length: 0

um DB seq length: 2000000000

-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

base : Issued Patents NA:*

- 1: /cgn2_6/prodata/2/ina/5A COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/6A COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/6B COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/6C COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

lt	o.	Score	Query Match	Length	ID	Description
1	2	663	21.4	2592	4	US-09-328-352-2604
2	2	649.4	20.9	1830121	4	US-09-557-884-1
3	3	649.4	20.9	1830121	4	US-09-643-990A-1
4	4	631.8	20.3	2628	4	US-09-107-532A-883
5	5	620.2	20.0	2574	4	US-08-887-534A-46
6	6	620.2	20.0	2574	4	US-09-527-431-46
7	7	594.8	19.2	2616	4	US-09-134-001C-1142
8	8	565.6	18.2	1230025	4	US-09-198-452A-1
9	9	546.8	17.6	2580	4	US-09-199-637A-280
10	10	546.8	17.6	2970	4	US-09-199-637A-272
11	11	546.8	17.6	42235	4	US-09-199-637A-1
12	12	545.2	17.6	2580	4	US-09-252-991A-4842
13	13	545.2	17.6	2970	4	US-09-252-991A-4897
14	14	495.4	16.0	3727	1	US-08-249-380-1
15	15	481.2	15.5	4403765	3	US-09-103-840A-2
16	16	481.2	15.5	4411529	3	US-09-103-840A-1
17	17	469.8	15.1	580073	4	US-08-545-528D-1
18	18	435	14.0	1674	4	US-09-724-623-18
19	19	376.6	12.1	2019	3	US-09-040-843-3
20	20	376.6	12.1	2019	4	US-09-621-855-3
21	21	376.6	12.1	2599	3	US-09-040-843-1
22	22	376.6	12.1	2599	4	US-09-621-855-1
23	23	370.2	11.9	2472	4	US-09-134-001C-1244
24	24	329.8	10.6	1677	4	US-09-199-637A-276
25	25	310.8	10.0	3274	4	US-09-221-017B-661
26	26	289.2	9.3	2640	4	US-09-252-991A-6033
27	27	283	9.1	2493	4	US-09-107-532A-2055

Sequence 2353, Ap
Sequence 8976, Ap
Sequence 684, App
Sequence 727, App
Sequence 123, App
Sequence 2033, Ap
Sequence 14, Appl
Sequence 11, Appl
Sequence 7, Appl
Sequence 5922, Ap
Sequence 168, App
Sequence 292, App
Sequence 4830, Ap
Sequence 294, App
Sequence 4859, Ap
Sequence 3366, Ap
Sequence 5908, Ap
Sequence 5965, Ap

ALIGNMENTS

RESULT 1

US-09-328-352-2604
; Sequence 2604, Application US/09128352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2604
; LENGTH: 2592
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2604

Query Match 21.4%; Score 663; DB 4; Length 2592;
Best Local Similarity 56.4%; Pred. No. 9.6e-182;
Matches 1277; Conservative 0; Mismatches 980; Indels 6; Gaps 2;

Qy	472	GGTCATCTCAATTTGGCTGTGACCACTGATGATGCTCTCTTGAAGATTCTCAATC	531
Db	310	GGCCATGAGTTTGTCTCACTGCTGTTTATGGCTTGGCAGAACTGGTGAACA	369
Qy	532	AGGATTTGTTGAACGAAGTCGGTGTAGCGACCGCGAGGCTAAAGTCTGAGTTGAGAAG	591
Db	370	AAAAATATTTTAACTGCGGTAGGTAACTCCCGACAGCTTACGCAAGTAATTGAAAT	429
Qy	592	CTTCGTCGGGAAGGAAGGAAGTTGAGAGTCTTCAGGGGACACAATTTTCAAGCT	651
Db	430	ATTCGAGTGTGACAAAGTCATGAGTAATAATCAGAACCAACGTCACCTTAAT	489
Qy	652	TTAAAGACTTTATGGAAGAGATTTGGTTGACCAAGCAGGGAAGCTTGCCTGTGATGGT	711
Db	490	AAATATACGATTGATTTTACCGAGCGGCTTTATCGGGGAACCTTATCCAGTGGTGGG	549
Qy	712	CGTGATGAGGATTTAGAGAGTCTGTGAGGATTTCTTTCGAGGAGAACGAGAACCAATCT	771
Db	550	CGTGATGAGGATTCGCGCTACCATTCAGTCTTGTACGCGCTACTAAAAATAACCCA	609
Qy	772	GTGCTTATTCGAGGACCGAGGATTTGTAACAGCTGTGTTGAGGTTTAGCACAAGG	831
Db	610	GTACTCATTCGTGAACCTGGGGTAGGTAAACCGCTTATTTGTAAGGTTTGGCACAACGT	669
Qy	832	ATTCTGAAGGAGATGTGCCCAACAGCTTCTTACTGATGTGAGATTAATTTCTGTTGACATG	891
Db	670	ATTGTCATTCGTGAAGTACCAGAGGCTTAAAGANTAAACGCTTTTATTCGTTAGATTTA	729

1810	Db	TTGTTTTTAGGACCAACAGGTGTTGGTTAAACCTGAGTTGACTAAAGCTTTAGCTAACTTC	1869
2029	QY	CTGTTTGATGATGAAGAAACCTCTTAGTTTCGATTTGATATGTCGGAATATATGGAACAACAC	2088
1870	Db	TTGTTTGACAGTGATGATGTCATGATTCGTATCCGATATGATGATGAAATTCATGAGAAACAT	1929
2089	QY	TCGTGCTCTCGCCTCAATGGGGCACCACACAGGCTATGTTGGTCACGAGGGAAGGTGGCAA	2148
1930	Db	TCGTGACCGCTTTAGTTTGTCGCCCTCCGGGTTATGTAGGTTTACGAAGAGGCGCGTGT	1989
2149	QY	CTTAACCTGAGCTGTGAGGAGGCCACCTTATTTGTGTCATACTCTTTGATGAAGTCGAGAAG	2208
1990	Db	TTAAACCGAGCTGTTGCCGCTAAACCATATAGCGTAGTGTGTTTGAAGAGGTTGAAAAA	2049
2209	QY	GCTCATGTTGCTGCTTTCAACACTCTGCTCCAAAGTTTGGATGATGCTGCAATTCAGAC	2268
2050	Db	GCGCATCCAGATGTCCTTCAATATCTTTGCTACAAGTGTAGACGATGACGCTTAAACCGAC	2109
2269	QY	GGGCAAGGCAGGACAGCTCGATTTCAAGAACTCGGTGATTAATCATGACATCAAAACCTTGGT	2328
2110	Db	TTCAAGAGTCTGTGTAGTGACCTTAAAAACAACGGTTATTGTTGATGACATCGAACTTGGGT	2169
2329	QY	GCTGAAACACCTCTTCGAGGCTTAACTCGGAAAGTAAACAATGGAAGTGCCTCCGGGACTGT	2388
2170	Db	TCACAGATGTAGCTGAA--CTTGGTGAAGGTGCACATGATGATGAATGCGCTACTATT	2226
2389	QY	GTGATCGGAGGTGAGGAAACAATTCAGACAGAGCTCTTCAACAGGCTTGACAGATT	2448
2227	Db	GTAATCAATCGGTAACTCAGCATTTCCGTCCAGAAATTTATTAAACGGATTTGATGATTG	2286
2449	QY	GTGGTGTTCGACCCCTTTTCATGACACAGTTGAGGAAAGTAGTCGGCTTCAAAATGAAA	2508
2287	Db	GTGATTTTCCATTCACTCAAAAAAGCACAGATTCGTGGCAATTCGCGATATTCAGTTGGAC	2346
2509	QY	GAGCTTGCTGTCGGGCTTCTGAAAGAGGAGTTGCTTTGGCAGTCACTGATGCTGCTTTG	2568
2347	Db	CGTTACGCTCAGCACTTGTTGATCGTGATATGAGTTTAACTGTAGATCAGAGTCGCAATT	2406
2569	QY	GACTATATCTTCGGCAGAGAGTTATGACCCGGTGTATGGTGTAGGCTTATAAGGAGATGG	2628
2407	Db	GACTTATTGATTTGACGCTGGTTTTGATCTCTGTATCGGAGCCGCTCCATTGAAACGTGCA	2466
2629	QY	ATGGAGAAGAAGGTGGTGACGAACTGTCAAAAGATGGTTGTGCGTGAGGAAATTCGATGAA	2688
2467	Db	AATTCAACAACAGGTTGAAATAACACTAGCTCAAAAAAATCTTGTCCAGGTGACTTTGTGGC	2526
2689	QY	AACTCCACTGTTTACATAGATCGAGCGCTGGTGATCTTGTT	2731
2527	Db	GGTGATACCAATTTTAGTTTAAAGGTGAAATGTGCTACTTAGTGT	2569

RESULT 2
US-09-557-884-1

; Patent No. 6506581

APPLICANT: Flei

; the

; * NUMBER OF SEQUENCES: 1

; ADDRESSEE: Human C

CITY: Rockville

COUNTRY: USA
ZIP: 20050

COMPUTER READABLE FORM: MEDTIM TYPE. 3 1 / 3

```
;
;
;
;
COMPUTER: Dell Per
OPERATING SYSTEM:
```


908379 TCGTCGTCATATTCAGTGAATCTTATTAGATGAAGTTGAAAAGCACACGAGATGATT 908438
 2226 CAACACTCTGCTCCAGTTTGGATGATGCTGATGACAGCGGCAAGGACGACAGT 2285
 908439 CAATATCTTATTACAGTCTGGATGATGCTGTTAATCTGATGTCAGGTGCTACTGT 908498
 2286 CGATTTACAGTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2345
 908499 GGACTTCGCTAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 908557
 2346 AGGCGCTAACTGCGAAGTAAACATGGAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCG 908557
 908558 -GTAATAAGAGCAAGCTATAGCAAAA-----TGAAGCCCTTAGTGTATTCATCCACT 908612
 908613 CCAACATTCGCGCAGAAATTCATCAACCGTATTGACGAAACCGGTTGATTCATCCACT 908672
 2466 TTCACATGACAGTGTGAGGAAGTAGCTCGGCTTCAATGAAGACGCTTGTCTCGGCT 2525
 908673 TGTAAAGAAATATTCGCTGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCG 908732
 2526 TGTGAAAGAGAGTGTCTTTGGCAGTCACTGATGCTGCTTTGGACTATATCTTGGCAGA 2585
 908733 GGAATCTGCTGATACGAATTTGGTGTATCCGATGCTTTATAGACTTCATTCGCGAAGT 908792
 2586 GAGTTATGACCGGCTGTATGCTGCTAGCGCTATAA 2620
 908793 GGGATACGACCAATTTATGTCGACGCTCAATTGA 908827

SUULT 3

-09-643-990A-1

Sequence 1, Application US/09643990A

Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann

Mark D. Adams

Owen White

Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of

the Haemophilus influenzae Rd Genome, Fragments

Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Query Match

Best Local Similarity 50.9%; Score 649.4; DB 4; Length 1830121;

Matches 1243; Conservative 0; Mismatches 946; Indels 6; Gaps 2;

QY	426	TCTTATTAAAGTCATTCGTCGCTCAAGCTCTCAGAGTCACGAGTGATPACTCAITT	485
Db	908639	TATTATCTACTTAATTTATGCGATAAAATTCACAGCAAAAACCAAGATAAATTTATTC	908698
QY	486	GGCTGTTGACCAAGTTGATTTATGGGTCTTCTTGAAGATTTCTCAATCAGGGATTGTTGAA	545
Db	908699	GAGCGAATTTGTTTTGTTTGGCAGCTTTAGAGACGAGAACGATCAGCGATATTTTGA	908758
QY	546	CGAAGTCGGTGTAGCGACGCGGAGGTTAAAGTCTCAGGTTGAGAAGCTTCGTGGGAAAGA	605
Db	908759	AAAGTCGGTGTGCAAAAAGAACAAATTTGCAAGCTATTTCAGCACATTAGAGGGGACA	908818
QY	606	AGGGAAGAAAGTTGAGAGTGTCTTCAGGGGACACAAATTTTCAAGCTTTAAAGACTTATGG	665
Db	908819	AAACGTGAACGATCAAAATGCAAGAAAGCAGAACAGCGCTTGAAAAATATACGATTGA	908878
QY	666	AAGAGATTGTTGAGCAAGCAGGGAAGCTTGTATCTCTGTGATGCTGATGAGGAGAT	725
Db	908879	TTTAAACCGCTCGTCGAGAAAGTGGCAACTTGTATCTGTAAATGCGGCTGATGAAGAAAT	908938
QY	726	TAGAAGAGTCGTGAGGATCTTTTCGAGGAGAACGAAGCAATCTGTGCTTATTGGAGA	785
Db	908939	TGCTCGAGCCATTCAAAGTATTACAAAGTCGTACCAAAATAACCTTGTTTAAATGGTGA	908998
QY	786	GCCAGAGTGTGTTAAACAGCTGTGTTGAGGTTTACGACAAAGGATTTGAAAGAGAGA	845
Db	908999	ACCAGGTGAGGAAACCGCGATTTGCGAAGGCTTGCGACAGCGCATCGTAAACGCGGA	907058
QY	846	TGTGCCCAACAGTCTTACTGATGTGAGATTAAATTTCTGTGACATGCTGCTGTAGTTC	905
Db	907059	AGTGCCAGAGGTTTGAAAAATAAACGTTGCTTTCATTAGATATGCGGCGGCTGTATGC	907118
QY	906	TGCTGCTTAATACCGAGAGGTTTGAAGAAAGGTTGAAATCTGTTTGAAGAAAGTTGA	965
Db	907119	TGGTGCGAAATATCGTGTGAAATTTGAAGAACGTTTAAAAAGCAGTACTCAATGAACCTTC	907178
QY	966	GGACGCTGAAGGCAAGTGATCTCTTTATTGATGAGATTCAITTTGTTCTTCTGTCGTCGG	1025
Db	907179	GAAAGAGAGAGGTGCGCGTTATCTCTTTATTGAGAAATTCATATGTCGCGCGCGG	907238
QY	1026	CAAACTGAAGGGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1085
Db	907239	TAAACCGCATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	907298
QY	1086	GCTTCGATGATGCTGCTACAGCTTGAAGATACAGAAATATGTTGAGAAAGATGC	1145
Db	907299	ATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	907358
QY	1146	TGCTTTGAGAGGAGGTTTCCAAAGTCTTATGTTGCGGAGCCAGGTTGCTGACACCAT	1205
Db	907359	CGCACTTGAACGCGGTTTCCAAAAAGTCTTGTGAGACGACCAAGTTGAGAGATACCAT	907418
QY	1206	TAGTATCTTTAGAGGACTCAAGAGAGATGATGAGGACATCATGTTGTCGCAATCCAGA	1265
Db	907419	TGCGATCTTACGTGGTTTGAAGAAAGCTTATGAAATTCATCATCATGCTGATATTACTGA	907478
QY	1266	CAGAGCTCTTATAAATGCTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1325
Db	907479	CCAGCAATTTGCTGCTGCGCAACGCTTTTTCATCGTGTATTTCCGATCGTCAGTTACC	907538

1326 GGATAAGCAATTGATTTGGTGTGATGAGGCTTGTGCAATGTGAGAGTCCAGCTGTAGT 1385
907539 AGATAAGCCATTGATTTGATCGATGAGAGCGCTTAGCATTTGATGGAATAGATTC 907598
1386 TCAACTGAGAGATTTGATAACCTTGAAGAGAGAGATGAGCTGGAATTTGAATTTCA 1445
907599 TAAACCTGAACCCCTTGTGCTGCTTGAAGCTGTATTTATATCAATTAATAATTTGGAACA 907658
1446 CGCCTTTGGAAGGAGAGAGATTAAGCCAGCAAGCTTCGACTTATAGAGGTGCGGAAGA 1505
907659 AGCGTTTACAAAAGAGAGAGAGCAAGCAAGCTTCGCAACCTTTAGAAATTTAGAGAAAGA 907718
1506 GCTTGTGATGACCTGAGAGACAAGCTTCAGCTCTTCAGCATGAATAACAGAAAGAGAGAA 1565
907719 ATTGGCTGAAAAGAGAGCTGAATACGCCGAATTTGAAGAAGTATGGAATCTGAAAAGC 907778
1566 GAGAAATGATGAGATTCGAAGGCTTTAAACAGAAAGAGAGAGCTCATGTTTCTTTTGA 1625
907779 AACGCTTTCTGGCTCTCAACATATTAAACAGAGTTAGATCTGCAAAACCGAACTAGA 907838
1626 GGAGGAGAGAGAGATATGACCTTCAAGAGCTGCTGATCTAAGATATGCGCAATTTCA 1685
907839 ACAAGCTGTGCGCGGGGATTTAGCGAAATGTCTGAGTTGCAATATGGCGCATCC 907898
1686 AGAAGTGAATCTGCAATTCGCCAACTTGAAGAACTTCTTCTGAAGAAATGTGATGCT 1745
907899 TGATCTTGAAGCACTTGAAGCACTGAACCCAGGAGGAAAGAAATGACGCTTT 907958
1746 CACAGAAACGTTGGGCGCTGAACATATGCTGAGGTTGTGAGCGGTGAGACAGGATTC 1805
907959 ACCTATCGCTCACAGATGAAGAAATCGCAGAGTGTCTTCTAAAGCCACAGGCAATCC 908018
1806 AGTGACGAGACTTGGCCAAATGAGAGAGAGAGTTGATTTGCTTGTGATGATTTGCA 1865
908019 TGTATCAAAATGATGGAAGGCGGAGAGAAAGAACTCTTGGATGGAAGATGAACATA 908078
1866 TAAGCGGGTTGTGGGACAGAACTCAACCGGTAAATGACGTTTCTGAGGCAATTTAAGGTC 1925
908079 TAAACGAGTATGTTGTCAGAGAGAGCGGTGATGCGGTAGCAACAGCATTTCTGCTAG 908138
1926 AAGGCGAGACTTGTGAGGACCAACAGCCCACTGGATCATTTCTTATTCCTTGGACCAAC 1985
908139 TGTGCGAGGCTTCTCGCATCTTAATCGCCCAATTTGCTTCTTCTTCTAGGCGCAAC 908198
1986 TGTGTTGCAAACTGAGCTGCCAAGCTCTTCTGAGAGAGCTGTTTGTGATGATGAAA 2045
908199 AGGTGTTGGAAACAGAGCTTTGCAAACTTTGGCTAAATTTCTGTTTGTATGATGAGA 908258
2046 CCTTTAGTTGCGATTTGATGTCGGAATATATGGAACCAACTCTGCTCTCGGCTCAT 2105
908259 TGGATGTTGCGTATTTGATGATGTCAGAGTTTATGGAACCAACAGGTTTCTCGTTAGT 908318
2106 TGGGGCACACAGGGTATGTTGTCACAGGAGGTGACAACTTAACCTAGCGCTGTGAG 2165
908319 TGTGCGGCTCCAGGCTATGTCGGCTATGAAGAAGCGGTTATTTAACTGAAGCTGTTG 908378
2166 GAGCGCACCTTATTTGTCATATCTTTGATGAGTGAAGAGGCTCATGTTGCTGCTT 2225
908379 TGTCTCTCAATTTAGTGAATCTTATAGTAAGTTGAAAGCAACGCAAGATGAT 908438
2226 CAACACTCTGCTCAAGTTTGGATGATGTCGATTCAGACAGCGGCGAGGAGGACAGT 2285
908439 CAATATCTTATTACAGTGTGATGATGTCGCTTTAACTGATGTCAGAGTCTGACTGT 908498
2286 CGATTTCAAGAACTCGGTGATATCATGATCAACCTTGGTGTGATGAACCTCTGTC 2345
908499 GGAATTCGGAACACTGTGTTGTTATGACCTTAACTTTGGGTTCTGATTTAATCCAAG- 908557
2346 AGGCTTAACCTGGAAGAGTACAAATGAAGTGGCGGAGTGTGATGCGGAGGTGAG 2405
908558 -GTAAATGAAGCAAGAGCTATACGAAA-----TGAAGCGCTTAGTGATGTCAGTGTAG 908612
2406 GAAACACTTCAGACAGAGCTCTTGAACAGGCTTTGACAGAGTTGTTGTTGCGACCCCT 2465

Db 908613 CCAACATTTCCGCCAGCAATTCATCAACCGTATTGCGAACCCTGATTCATCCACT 908672
QY 2466 TTCACATGACCAAGTTGAGAAAGTAGCTGGCTTCAATGAAGACCGTTGCTGTCCGCT 2525
Db 908673 TGGTAAAGAAATATCCGTGCGATTGCAAGTATCCAATTAGAACGCTTAGCAAAACGTAT 908732
QY 2526 TGCTGAAGAGAGCTTCTTTGGCAGTCACTGATGCTGCTTTGCACTATATCTTGGCAGA 2585
Db 908733 GGAACACTGCTGATGACGAATTTGGTGTACCGATGCTTTATTAGACTTCATTGGCGAAGT 908792
QY 2586 GAGTTATGACCCGCTGATGCTGCTAGGCTTATAA 2620
Db 908793 GGGATACGACCAATTTATGCTGACGCTCCATTGA 908827

RESULT 4
US-09-107-532A-883
; Sequence 883, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION NUMBER: 60/085,598
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 883:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2628 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...2628
; SEQUENCE DESCRIPTION: SEQ ID NO: 883:
US-09-107-532A-883

Query Match 20.3%; Score 631.8; DB 4; Length 2628;
Best Local Similarity 57.3%; Pred. No. 1.1e-172;
Matches 1183; Conservative 0; Mismatches 877; Indels 6; Gaps 2;

Db	1622	CCGAAATGAGATTGCTCAAGTCGTTGGCAGATTAAACGGGTATCCCGGTACGAAACTTG	1681
QY	1820	GC AAAATGAGAAGGAGAGGTGTGATCGTCTTGCTGATAGGTTGCATAAGCGGGTGTGG	1879
Db	1682	TAGAAGGTGACGAGAGAAAACCTAATTAAATTGAATGAAAACATTCATTAACGAGTGA	1741
QY	1880	GACAGATCAAGCGGTAAATGCGATTTCTGAGGCAATCTTAAGTCAAGGCGAGACTTG	1939
Db	1742	GTC AAGATGAGCAGTGTGATCGCGTCAAGTCACGCGGTAAATCGTTTACGCGGTGGA	1801
QY	1940	GTAGGCGACAAACGCCAACTCGATCAATCTTATTCCTTGGACCAACTGCTGTGGC	1999
Db	1802	AAGACCCAAATCGTCCCTAGTGTCTTTCCTTTTCCPAGGACCAACTGCTGTGG	1861
QY	2000	CTGAGCTCGCCAAAGCTCTTGCTGAGCAGCTGTGTTGATGATGAAAACTCTTAGTTC	2059
Db	1862	CCGAATTAGCGAAGCGTTAGCTGAGAACTTGTTTGA CT CAGAAGATCATATGGTAC	1921
QY	2060	TTGATATGTCGGAATATATGAGAACAAACATCTGTCTCTCGCCCTCATTTGGGGCA	2119
Db	1922	TCGATATGAGTGAATATATGCGAAAAACATGCTGTGTCTCGCTAGTCGGAGCTCT	1981
QY	2120	GGTATGTTGTCACGAGGAAGTGGCAACCTAACTGAGGCTGTGAGGAGCGGACCTTA	2179
Db	1982	GCTATGATAGTTACGAAGAAGTGGTCAATTAAACGGAAGCTGTGACGAAAAATC	2041
QY	2180	GTGTCAATCTTTTGATGAAGTGGAGAAGGCTCATGTTGCTGTCTTCAACACATCT	2239
Db	2042	CAATCGTCTTTGGATGAATCGAAAAAGCGCATCCCGATGCTTCAATATCTGT	2101
QY	2240	AAGTTTGGATGATGGTGCATTTGACAGACGGGCAAGCAGGACGATGCTGATTCAG	2299
Db	2102	AAGTCTCAGATGACGGACGCTTGACTGATTTCAAAAGGGCGAGTAGTAGTTTCA	2161
QY	2300	CGGTGATAATCATGACATCAAACTTGTTGCTGCAACCTCTTGCAGGGCTAACTGG	2359
Db	2162	CGGTTCTGATCATGACGCAATATCGTTTCA CAGTTACTGCTGGAAGGCGTACTG	2221
QY	2360	AAGTAACAAT- --GGAATGGCCCGGACGTGTGTGATGGGAGGTGAGAAACATTC	2416
Db	2222	ATGNAACGATTCAGAAAGCAGTGGCTGAACAAGTCAACAGCTATTACGTGGAACT	2281
QY	2417	GACCAGACTCTTTGAACAGGCTTGACGAGATTGTGTGTTTCGACCCCTTTCATG	2476
Db	2282	AACCGAATTTCTTGAACCGCAATCGACGACACGATTTTATTTACACCATTTG	2341
QY	2477	AGTTTGAGAAAGTAGCTCGGCTTCAATGAAGACGTTGCTGTCGGGCTTGTG	2536
Db	2342	ACGTGAAGGAATCGTTTGACAAATCGTGGCGCAATGGCAACAAGCTCTGGAAC	2401
QY	2537	GAGTTGCTTTGGCAGTCACTGATGCTGCTTTGGACTATATCTTGGCAGAGTATG	2596
Db	2402	AAATCCTGCTGACAACTAGTGATGAAGCCAAAAACGTGGATTGCTGAAAAATG	2461
QY	2597	CGGTGTATGCTGTAGGCTTATGAAGAGATGGAATGGAAGAAGGTGTGTGACAG	2656
Db	2462	CAGCATATGCTGTCAAGACCATTTGAACGATTTTATCACTAAGGAAGTAGAAA	2521
QY	2657	CAAGATGTTGTGCGTGAAGGAAATCGATGAAAACTCCACTGTTTACATAGATG	2716
Db	2522	CTAAAGAAATCGTTGACGACATGTTATGCGCTTAATCAAAAGTAAACGATTA	2581
QY	2717	CTGGTGATCTTGTTACCGGGTAGAA	2742
Db	2582	ATGCAAACTTCAATTTCAAAAACAGAA	2607

RESULT 5
US-08-887-534A-46
; Sequence 46, Application US/08887534A
; Patent No. 6455323
; GENERAL INFORMATION:

APPLICANT: Holden, David W.
 TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 STREET: 233 South Wacker Drive/6300 Sears Tower
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/887,534A
 FILING DATE:
 CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien
 REGISTRATION NUMBER: 33,547
 REFERENCE/DOCKET NUMBER: 28341/33996
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 474-6300
 TELEFAX: (312) 474-0448
 TELEX: (312) 474-6600

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:
 LENGTH: 2574 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "DNA (genomic) (p10b30)"
 FEATURE:

NAME/KEY: CDS

LOCATION: 1..2571

08-887-534A-46

Query Match 20.0%; Score 620.2; DB 4; Length 2574;
 Best Local Similarity 54.7%; Pred. No. 2.4e-169;
 Matches 1278; Conservative 0; Mismatches 1053; Indels 6; Gaps 2;
 352 AGAGTGATCAATCAAGCCTTGAAGAGCTTCTTCAATCTCTCCACCTGATGATTT 411
 184 ACAGATATCAATCAGGCATTAATCGTTTACCGCAGGTTGAAGGTACTGGTGGTATGTC 243
 412 CCAGCGAGTTCAGTCTTATTAAGGTCAATTCGTCTGCTCAAGCTGCTCAGAGTCAGGA 471
 244 CAGCCATCACAGGATCTGGTGGCGCTTCTTAATCTTTGCGACAACGTCGCGCAAAACGT 303
 472 GGTGATCTCAATTTGGCTGTTTACAGGTTGATTATGGGTCTTTTGAAGATTCCTCAATC 531
 304 GGTGATACTTATCTGCTCAGAACTGTTTCTGGCGGCACTTGAATCTCGCGGACCC 363
 532 AGG---GATTTGTTGAACAGAGTCGTGTAGCAGCGGAGGGTAAAGTCTGAGGTTGAG 588
 364 GTGGCGGACATCTCTGAAAGCAGCAGGGGCGCACCCGCAACATTAATCAAGCGATTGAA 423
 589 AAGCTTCGTGGGAAGAGGGAAGAGCTTCAGAGTGTCTTCAAGGGAACAAATTTCAA 648
 424 CAATGCTGGAGGTGAAGCGTGAACGATCAAGGTGCTGAAGCAACCAACGTCAGGCTTG 483
 649 GCTTTAAAGACTTATGAAGAGATTGTTGTTGAGCAAGCAGGGAAGCTTGATCTCTGATTT 708
 484 AAAAATAATACCATCGACTTACCGAAGCAGCCGGAACAGGCAACTCGATCCGCTGATT 543
 709 GGTCTGATGAGGAGATTAGAGAGTCTGTAGGATTTCTTCGAGGAGAACGAACAAT 768
 544 GGTCTGATGAAGAAATTCGCGGTACCATTCAGGTGCTGCAACGCTGCTACTTAAATAAC 603
 769 CTTGCTGCTTATTGGAGAGCCAGGAGTTGGTAAACAGCTGTGGTTGAAGGTTTACGCAAC 828

604 CCGGTACTGATTTGGTGAACCCCGGCTCGTAAACTGCATCTGTTGAAGGCTCTGGCGCAG 663
 829 AGGATTTGAAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTCTGTTGAC 888
 664 CTTATTATCAACCGCGAGTCCGGAAGGTTGAAGCGCCGCGGTACTGGCGCTGGAT 723
 889 ATGGGTGCTTGTAGTTGCTGCTGCTAAATACCGAGGAGGTTTGAAGAGGTTGAAATCT 948
 724 ATGGCGCGCTGTGGCTGGCGGCGGAATATATCGCGTGAATTTGAAGACGTTTAAAGGC 783
 949 GTTTTGAAGAAAGTTGAGGACGCTGAAGGCAAAAGTATTCTCTTTATTGATGAGATCAT 1008
 784 GTGCTTAACGATCTTGCCAAACAGAGGACCACTCATCTTATTINTCGACGAATTTACAT 843
 1009 TTGGTTCTTGTGCTGGCAAACTGAAGGTCGATGATGATGATGATGATGATGATGATGATG 1068
 844 ACCATGCTGGCGCGGTAAGCGGATGGCGCAATGAGCGCCGGAACATGCTGAAACCG 903
 1069 ATGTTAGCTAGAGGCGCAGCTTTCGATGCTTGGTCTACAGGCTTGAAGATACAGGAAA 1128
 904 CGGCTGGCGGCTGTGATTTGCACTGCTGAGTGGCCAGGCTTGAAGCAATATCGCCAG 963
 1129 TATGTTGAGAAAGATGCTGCTTTGAGAGGAGGTTCCAAACAGTCTATGTTGCGGAGCCA 1188
 964 TACATTTGAAAAGATGCTGCTGCTGGAACGTCGTTTCCAGAAAGTGTGTTGCGGAGCCT 1023
 1189 AGTGTGCTGACACCATTAATGCTTTAGAGGACTCAAGAGAAAGTATGAGGAGATCAT 1248
 1024 TCTGTTGAAGATACCATTTGCGGCTGAGGCTGGAAGAGCTTACGAATTTGCAACAC 1083
 1249 GGTGTGCAATCCAAAGACAGCTCTTATAAATGCTGCTCAGCTGTCTGCTGCTGATATA 1308
 1084 CATGTGCAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1143
 1309 ACTGCTGCGCAATTTACCGGATAAAGCAATTTGATGATGATGATGATGATGATGATGATG 1368
 1144 GCTGACCGCTCAGCTGCGGATAAAGCAATGACCTGATGATGATGATGATGATGATGATG 1203
 1369 AGAGTCCAGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1428
 1204 CGTATGCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1263
 1429 CTGAAATTCGAATTCAGCGCTTGGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488
 1264 CTCAACTGGAACCAACAGCGCTTAAAGAGTCTGATGAGGCTGATGAGGCTGATGAGG 1323
 1489 ATAGAGTGGGAAAGAGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1548
 1324 GATATGCTCAACGAGAACTGAGCGCAACAAAGACGTCAGTACTCCGAGTTAGNAGAGAG 1383
 1549 TACAGAAAGGAG 1608
 1384 TGAAGAGCAG 1443
 1609 CTCAATGTTTCTTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1668
 1444 GCGAAATCGCTATTGAAACAGGCTGCGCTGTTGGGGGAGCTGGCGCGGATGTTCTGAA 1503
 1669 AGATATGGCGCAATTCAGAGTGGAACTCTGCAATTTGCCCAACTTTGAAGGAACCTTTCT 1728
 1504 CAATACGCAAAATCCCGAACTGGAAGAGCAACTGGAAGCGCAACGCAAGCTCGAAGGC 1563
 1729 GAGAGAAATGATGCTCAGAGAAACGTTGGGCGCTGAAACATGATGCTGAGGTTGTGAGC 1788
 1564 AAAATATGCTGCTGTTGCTGTAATTAAGTGAACGCGCGCAAAATGCTGAAAGTGTGCG 1623
 1789 CGTTGACAGAGGATTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1848
 1624 CGTTGAGCGGAGATTCGCGGTTTCTGCTGATGATGATGATGATGATGATGATGATGATG 1683
 1849 CTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1908

1684 ATGGAGCAAGAACTGCACCATCGCGTAATTGGTCAGAAAGAAAGCGGTTGATCGGTATCT 1743
1909 GAGGCAATTTCTAAGGTCAAGGCGAGGACTTGGTAGGCAACAGCAACCACTGATCATCTC 1968
1744 AACGCTATTTCGTAGCGGTGGGGCTGGCGGATCCAAATCGCCGATTTGGTTTCATTC 1803
1969 TTATTCCTTGGACCAACTGGTGTGGCAAACTGAGCTGCCAAGGCTCTTCTGAGCAG 2028
1804 CTGTTCTCGGCCCAACTGGTGTGGGGAACAGAGCTTTGTAAAGCGCTGCGCAACTTT 1863
2029 CTGTTTGTATGATGAAGAACTCTTAGTTCGGATTGATATCGGATATATGGAACAAAC 2088
1864 ATGTTTGTATGAGCAAGAGCGATGGTCCGATTCGATATGTCGAGTTTATGGAGAAAC 1923
2089 TCTGTCTCTCGCTCATTTGGGCAACACAGAGGATGTTGGTTCAGAGAAAGGTGGCAA 2148
1924 TCGGTGTCTCGTTTGGTGTGGTCCGCTCGGATATGTCGGTTATGAAGAGGTGGCTAC 1983
2149 CTAACCTGAGGCTGTGAGGAGGCACTTATTTGTGTCTATCTTTGATGAAGTGGAGAG 2208
1984 CTGACCGAAGCGGTGCGTCTCGGATTCGCTCATCTCTGCTGGATGAAGTGGAAAA 2043
2209 GCTCATGTTGCTGCTCTCAACACTCTGCTCCAAAGTTTGGATGATGTCGATTCGACGAC 2268
2044 GCGCATCCGATGCTTCAACATCTGTTGAGGACTGCGATGATGGGCGTCTGACTGAC 2103
2269 GGGCAAGCAGGACAGTCCGATTTCAAGAACTCGGTGATATCATCATCAAACTTGGT 2328
2104 GGGCAAGGAGAGCGTCCGACTTCCGTAATACGGTCTGCTCATTTGACCTCTAACTCGGT 2163
2329 GCTGAACACCTCTTTCAGGCTTAACCTGGAAAGTAAACATGGAAGTGGCCCGGACTGT 2388
2164 TCCGATCTGATTCAAGAACGCTTCGGTGAACGTAATATGCGCATG---AAAGAGCTG 2220
2389 GTGATGCGGAGGTGAGGAAACACTTTCAGACAGAGCTCTTGAACAGGCTTTCAGGAGAT 2448
2221 GTGCTCGGTGTGTAAGCAATATCTTCGTCGGAATTCATTACCGTATCATGATGAAGT 2280
2449 GTGGTGTGCAACCCCTTTCATGACAGCAGTTCAGTGAAGAGTTCGCTTCAATGAAA 2508
2281 GTGGTCTTCCATCCGCTGGTGAACAGCAGCATGCTCGATTCGAGATTCAGTTGAAA 2340
2509 GACGTTGCTGTCGCGCTTTCGAAAGAGAGTTCGTTGGCAGTCACTGATGCTGCTTTG 2568
2341 CGTCTGTACAAACGCTTCGAAAGACGTTGTTATGAATCCAAATTCGACGAGCGCTG 2400
2569 GACTATATCTTGGCAGAGATTATCACCGGCTGTATGCTGCTAGGCTTATGAGGAGATG 2628
2401 AAACCTGCTGAGCAGAGAGGTTACGATCCGCTCTATGTTGCACTCTCTGAAACGTGCA 2460
2629 ATGAGAGAGAGTGGTGACAGAACTGTCAAGATGTTGTCGCTGAGGAAATCGAT 2685
2461 ATTCAGCAGCAGATCGAAACCCGCTGSCACAGCAAAATATGCTGCTGCTGAATTGGTT 2517

RESULT 6

-09-527-431-46

Sequence 46, Application US/09527431

Patent No. 6485899

GENERAL INFORMATION:

APPLICANT: Holden, David W.

TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive/6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/527,431
APPLICATION NUMBER: US/09/527,431
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: US/08/887,534
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 28341/33996
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: (312) 474-6600
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 2574 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA (genomic) (p10b30)"
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2571
US-09-527-431-46

Query Match 20.0%; Score 620.2; DB 4; Length 2574;

Best Local Similarity 54.7%; Pred. No. 2.4e-169;

Matches 1278; Conservative 0; Mismatches 1053; Indels 6; Gaps 2;

QY 352 AGAGTATCATCAAGCCTTGAAGAGCTTCCTTCAAAATCTCCACCTGATGATTT 411
DB 184 ACAGATATCAATCAGGCATTAATCGTTTACCGCAGGTTGAAGGTACTGTTGATGTC 243
QY 412 CCAGCGAGTCTTAGTCTTATTAAGGTTCATTGCTGCTGCTCAAGCTCTCAGAAGTCAG 471
DB 244 CAGCCATCAGAGTCTGTCGCGGTTCTTAACTTTGGGCAACGTCGCGCAAAAAGT 303
QY 472 GTGATATCATTTGGCTGTTCACCAAGTTGATTTGGTCTTCTTGAAGATTTCTCAAAATC 531
DB 304 GGTGATAACTTTATCTGTCAGACTGTTCTGTTCTGCGGCACCTTGGTCTCGCGGACC 363
QY 532 AGG---GATTGTTGAAAGAGTCTGCTGAGTTCAGCAGCGAGGTAAGTCTGAGTTCAG 588
DB 364 GTGGCCGACATCTGAAAGCAGCAGCGGCGGACCCCGCCAACTTACTCAAGCGATTGAA 423
QY 589 AAGCTTCGTGGGAAAGAGGAAAGTTGAGAGTTCAGAGGCTTCAGGGGACACAAATTTCAA 648
DB 424 CAAATCGCTGGAGGTGAAGCGTGAACGATCAAGTGTCTGAGACCAACGTCAGGCTTTG 483
QY 649 GCTTTAAGACTTATGGAAGATTTGGTTGAGCAAGCAGGGAAGCTTGTCTCTGTGATT 708
DB 484 AAAAAATATACCATCGACTTACCAGAACGAGCGGAAACAGGCGCAACTCGATCCGCTGAT 543
QY 709 GGTCTGTGATGAGGAGATTAAGAGTCTGAGGATTTCTTCGAGGAGAACGAGACAAAT 768
DB 544 GGTCTGTGATGAAGAAATTCGCGTACCATTCAGGTGCTGCAACCGTCTACTAAAAATAC 603
QY 769 CCTGTGCTTATTGGAGAGCCAGGAGTTCGTTAAAAACAGCTGTGGTTGAAGGTTTAGCAAA 828
DB 604 CCGTACTGATTGGTGAACCCCGGCTCGGTAAGTTCGATTCGTTGAAGTCTGCGCGAG 663
QY 829 AGGATGCAAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTCTGTGAC 888
DB 664 CGTATTATCAACGCGAAGTGTCCGGAAGGTTTGAAGGCGCCCGGCTACTGGCGCTGAT 723
QY 889 ATGGGTGCGTGTGTTGCTGCTGCTAAATACCGAGGAGTTCGAGGAGGTTTGAATCT 948
DB 724 ATGGGCGCGCTGCTGGCTGGGCGGAAATATTCGCGGTGAGTTTGAAGAACGTTTAAAGGC 783

2029	QTGTTTGATGATGAAGAAACCTCTTAGTTCCGATATGTCGGAAATATATGGAAACAACAC	2088
1864	ATGTTTGATAGCGACAGACGGATGGTCCGTATCCGATATGTCGAGTTTATGGAGAAACAC	1923
2089	TCTGTCCTCTCGCCTCATTTGGGGCACACACAGGGTATGTTGGTCACGAGGAAGGTGGACAA	2148
1924	TGGGTGTCCTCGTTTGGTTCGGCCTCCGGGATATGTCGGTTATGAGAAAGGTGGCTAC	1983
2149	CTAACTGAGGGCTGTGAGGAGGGACCTTATTTGTCATATCTTTGATGAAGTGGAGAAG	2208
1984	CTGACCGAAGCGGTGGTCGTCTCGGTATTCGGTCATCTCTGCTGGATGAAGTGGAAAAA	2043
2209	GCTCATGTTGCTCTCTTCAACACTCTGCTCCAAAGTTTTCGATCATCGTCGATTTGCACAGAC	2268
2044	GCGCATCCGGATGTCCTTCAACATCTCTGTCAGGTACTCGATATGGCGCTTCGACTGCAC	2103
2269	GGCGAAGGCGAGACAGCTCGATTTTCAGAACTCGTGATATCATATGATCATCAAACTTTGGT	2328
2104	GGCGAAGGAGAAACGGTCGACTTCGGTATACGGTCGTCAATTATGACCTCTAACCTCGGT	2163
2329	GCTGAACAACCTCTTCGAGGGCTAACTGGGAAAGTAAACAATGGAAGTGGCCCCGGGACTGT	2388
2164	TCCGATCTGATTCAGGAACGCTTCGGTGAACCTGGATATCGCGCACATG---AAGAGAGCTG	2220
2389	GTGATGCGGGAGGTGAGGAAAACACTTTCAGACCAGAGCTCTTTGAAACAGGCTTGACAGAGATT	2448
2221	GTGCTCGGTGTGTATAGCCATTAACCTTCGTCGGAATTCATTAACCGTATCGATGAAGTG	2280
2449	GTGGTGTTCGACCCCTTTTCACATGACCAGTTGTGAGGAAAAGTAGCTCGGCTTCAAAATGAAA	2508
2281	GTGGTCTTCCATCCGCTGGGTGAAACAGCACATTTGCCCTCGATTGCGCAGATTCAGTTTGAAA	2340
2509	GAGGTGTCTGTCCGGCTTGCTGAAAGAGAGAGTTGCTTTTGGCAGTCACTGATGCTGCTTTG	2568
2341	CGTCTGTATAAAACGCTCTGGAAGAAACGTGGTTATGAATCCACATTTCTGACGAGGCGGCTG	2400
2569	GACTATATCTTCGCACAGAGTTTATGACCCGGGTGATGTTGCTAGGCTTATTAAGGAGATGG	2628
2401	AAACTGCTCAGCGAGNAACGGTTACGATCCGGTCTATGTTGTCACGTCCTCTGAAACGTGCA	2460
2629	ATGGAGAAGAGGTGGTGTGACGAACCTGTCAAGAATGGTTGTGCGTGAGGAAAATCGAT	2685
2461	ATTTCAGCACGATCGAAAACCGGCTGGCACAGCAAAATACTGTCTGGTAAATGGTT	2517

US-09-134-001C-1142

; PATENT NO. 6380370
: GENERAL INFORMATION:

APPLICANT: LYNN DOUGLAS
; TITLE OF INVENTION:

; FILE REFERENCE: GTC

CURRENT FILING DATE

PRIOR FILING DATE:

; PRIOR FILING DATE:

; SEQ ID NO 1142

TYPE: DNA

US-09-134-001C-1142

Query Match
Best Local Similarity

Matches 1184; Converse

QY 637 GACACAFAR

1753	1606	1813	1666	1873	1726	1933	1786	1993	1846	2053	1906	2113	1966	2173	2026	2233	2086	2293	2146	2353	2206	2410	2266	2470	2326	2530	2386	2590	2446	2650	2506		
QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	Db	QY	
1753	1606	1813	1666	1873	1726	1933	1786	1993	1846	2053	1906	2113	1966	2173	2026	2233	2086	2293	2146	2353	2206	2410	2266	2470	2326	2530	2386	2590	2446	2650	2506		
AACTTGGCCCTTGACACACANTGCTCAGGTTGTGAGCCGTTTGGACAGGGATTCACAGTGACG	GTCTGTTTTCAGATGAGAAATAGCGCACATAGTTAGTCAATGAGCAGGTATACCTGTATCT	AGACTTGGCCAAAATGAGAAGGAGAGGTTGATTTGGTCTTCTGATAGTTGGTCATAAAGCGG	AAACTTGTGGAAACAGAAAGAGAAAAGCTATTAAAGTTTAAAGTGACATCTTTACACAAACGT	GTGTGGGACAGAACTCAAGCGGTAATGCAGTTTCTGAGGCAATTCCTAAGGTCGAAGGGCA	GTATGTAGGTCAGATAAAGCAGTTGATTTAGTATACAGACGAGTAGTTAGAGCACGTGCT	GGACTTGGTAGGGCACACACAGCCAACTGGATCATCTTATTTCTTCTGGACCAACTGCTGTT	GGAAATTAAGATCCGAAATAGACCAAATCGGAAGTTTCTTATTTCTTAGGACCTCATCTGGAGTA	GGCAAACTGAGCTCCCAAGGCTCTTGTCTGACGACGTGTTTGCATGATGAAAACCTCTTA	GGTAAACCTGAATTAGCAAAATCGCTTGTCTTCATCACTTTTTCGATTTCTGAAAACATATG	GTTCGATTTGATATGTCGGAATATATGGAACAAACACTCTGTCTCTCGCTCCATTTGGGGCA	ATTAGAAATTTGATATGAGCAGATATATGGAACAACTGCTGTATCACGTTTAAATGGTGCA	CCACCAAGGTPATGTTGGTCAACGAGGAAGGTGGAACAATACTGAGGCTGTGAGGAGCGCA	CTTCCAGGTTATGTAGGTCACGATGAAGGTGGTCAATTAATCTGAAGCAGTTAGACGTAAT	CTTATTGTCTCATCTCTTTGTATGAAGTGGAGGAGGCTCATGTGCTGTCTTCAACACT	CCATCTCAGTTATTTTGTGTAGACGAAGTTGAAAAGCACATAGCGATGTTTTTAATGTA	CTGCTCCAACTTTTGGATGATGTFPGATGACAGCGGGCAAGGACGACAGTCGATTTTC	TTACTTCAAATACTAGATGAAGTCGTCTTACGGAATCTTAAAGGTAGAAGTGTGTGACTTT	AGGAATCTCGGTGATATCATGACATCAAACTTTGGTGCTCAACACCTCTTTCAGAGGCTA	AAAAATACCAATTATCATCATGACTAGTAATATTGTTTCACAAGTATTACTTTGAAAATGTA	A---CTGGGAAAAGTAAACAAATGGAAGTGGCCCGGACCTGTGTGATGCGGGAGGTGAGGAAA	AAAGATGCTCGTGAAATTAGTGTGATGATACAGAGAAAGCAGTTATGGACAGTCTACATGCA	CACTTCAGACACAGACTCTTTGAAACAGCGCTTGAACGAGATTGGTGTGTTTCGACCCCTTCA	TACTTCAACCTGAAATATTAAATTCGATCGATGACATCGGTATTATTTAAACCAATTATCA	CATGACCAGTTTGAGAAAAGTAGCTCGGCTTCAAAATGAAAAGCAGTTGTGCTCGCGCTTGCT	GTTCATGATATGAGTATGATTTGTGTAGATAAAATTTTAAACAAATTAATATGAGATTATTA	GAAGAGAGAGTTGCTTTTGGCAGTCACTGTATGCTCTCTTTGGACTATATCTTGGCAGAGAT	GATCAACATATCTCAAATTGAAGTGACAGAGAAGCGGAAAAAATCGCTAGGTGAGAGAGCG	TATCACCGGTTGATGTTGCTTAGCGCTTATAGGAGATGAGTGGAGAGAAGGTTGGTGACA	GATCAACATATCTCAAATTGAAGTGACAGAGAAGCGGAAAAAATCGCTAGGTGAGAGAGCG	TATCACCGGTTGATGTTGCTTAGCGCTTATAGGAGATGAGTGGAGAGAAGGTTGGTGACA	TATGAAACCAAAITTTGGTGCAAGACCAATTTAAACGCTTTGTTTCAACGACAAATAGAACT	GAACTGTCAAGATGTTGTGTCGTGAGGAAATCGATGAAAACCTCAACTGTGTTTATCATGAT	CCAAATGCACTATGATGATTTAAAGAAAGTCTACTGAGAGGTACAATAATTTAAAGTAGAT

US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifais, R.
; TITLE OF INVENTION: Chlamydia pneum

TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention of infection
 TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention of infection
 REFERENCE: 9710-003-999
 CURRENT APPLICATION NUMBER: US/09/198,452A
 FILING DATE: 1998-11-24
 NUMBER OF SEQ ID NOS: 6849
 Q ID NO 1
 LENGTH: 1230025
 TYPE: DNA
 ORGANISM: Chlamydia pneumoniae
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (1)..(15000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (15001)..(30000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (30001)..(45000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (45001)..(60000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (60001)..(75000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (75001)..(90000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (90001)..(105000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (105001)..(120000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (120001)..(135000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (135001)..(150000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (150001)..(165000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (165001)..(180000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (180001)..(195000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (195001)..(210000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (210001)..(225000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (225001)..(240000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (240001)..(255000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (255001)..(270000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (270001)..(285000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (285001)..(300000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (300001)..(315000)
 OTHER INFORMATION: n=a or c or g or t

OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (315001)..(330000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (330001)..(345000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (345001)..(360000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (360001)..(375000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (375001)..(390000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (390001)..(405000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (405001)..(420000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (420001)..(435000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (435001)..(450000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (450001)..(465000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (465001)..(480000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (480001)..(495000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (495001)..(510000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (510001)..(525000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (525001)..(540000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (540001)..(555000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (555001)..(570000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (570001)..(585000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (585001)..(600000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (600001)..(615000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (615001)..(630000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (630001)..(645000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (645001)..(660000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (660001)..(675000)
 OTHER INFORMATION: n=a or c or g or t

2005 CTCGCCAGGCTCTTGCTGAGCAGCTGTTTGTATGATGAACAACTCTTAGTTCGGATTGAT 2064
190175 CTTGCAAAAGCTCTTGAGATCTCTTTTCAATAAAGAGAACTATGTCGCTTCGAT 190234
2065 ATGTCGAATATATGAAACAACACTCTGCTCTCTGCTCTATTTGGGCGACACACAGGATAT 2124
190235 ATGTCGAGATATATGAAAGAGCAATTCATTTCCAAGCTTATAGGATCTTCTCCAGGATAT 190294
2125 GTTGGTCCAGCAAGAGTGCACAACTAACTAGGCTGTGAGGAGGCGACTTATTGTGTC 2184
190295 GTGGTTATGAGAGAGTGGAGTCTTTCTGAGGCTCTTCGACAGCTCCCTATTTCAGTA 190354
2185 ATACTCTTTGATGAAGTGGAGAGGCTCATGTTGCTCTTCAACACTCTGCTCCAAGTT 2244
190355 GTTCTCTTTGATGAGATAGAGAAAGCAGATAAGGAAAGTTCTTAATATCTTTTACAGGTT 190414
2245 TTGGATGATGCTGATGACAGACGGCGGACGAGCAGACAGTCAATTTAGGAACCTCGGTG 2304
190415 TTTGATGATGGGATCTTACGGATGGGAAACGCAAAAGTAAATTTGTAATAATGCTTTG 190474
2305 ATAAATCATGACATCAAACTTGGTGTGAACACCT-----CCTTCAGGGCTTAAC 2355
190475 TTTATCATGACATCAATATAGGTTCTCCAGAACTTGCAGATTATTTGTTCAAAAAGGA 190534
2356 GGGAAAGTAAATGAAGTGGCCCGGCACTGTGTGATGCGGAGGTGAGGAAACACTTC 2415
190535 AGTGAGCTTACGAAAGAGCGATTCTTTCTGTAGTCTCTCCAGTATTGAAAGATACCTG 190594
2416 AGACGAGCTCTTGAACAGCTTGAAGAGATTGAGGATGTTGGTGTTCAGACCTTTTCAATGAC 2475
190595 AGCCCTGGAATTTATGAACCGAAATGATGATGAGATACCTCTTTGTTCCATTAACGAAAGAA 190654
2476 CAGTTGAGGAAAGTAGCTCGGCTTCAAATGAAGAGCTTGTCTCGGCTTGTCTGAAGA 2535
190655 GATATCGTGAATATGTCATCAATGCGAGGATTTGCCAGAGATTAAGGCAAG 190714
2536 GAGTTGCTTTGGCAGTCACTGATGCTGTTGGATATATCTTGGCAGAGATTATGAC 2595
190715 CGATCAATTTATCTTGGGATGATCTGTAATATTTCTTAGTGAACAGGTTATGAC 190774
2596 CCGGTATGTTGCTAGGCTTATAGGAGATGATGAGAGAGAGTGTGACAGACTG 2655
190775 AGTGCTTTCGGAGCCGCCCTTTAAACAGTTTATCCAAACAAAAGTTGTGATCTTGTCTT 190834
2656 TCAAG 2661
190835 TCTAAG 190840

LT 9
9-199-637A-280
quence 280, Application US/09199637A
tent No. 6355411
NERAL INFORMATION:
PPPLICANT: Ausubel, Frederick
PPPLICANT: Goodman, Howard M.
PPPLICANT: Rahme, Laurence G.
PPPLICANT: Mahajan-Miklos, Shalina
PPPLICANT: Tan, Man-Wah
PPPLICANT: Cao, Hui
PPPLICANT: Drenkard, Eliana
PPPLICANT: Tsongalis, John
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
FILE REFERENCE: 00786/361002
CURRENT APPLICATION NUMBER: US/09/199,637A
CURRENT FILING DATE: 1998-11-25
PRIOR APPLICATION NUMBER: 60/066,517
PRIOR FILING DATE: 1997-11-25
NUMBER OF SEQ ID NOS: 437
SOFTWARE: FastSeq for Windows Version 4.0
O I D NO 280
LENGTH: 2580

TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-280

Query Match
Best Local Similarity 53.8%; Score 546.8; DB 4; Length 2580;
Matches 1158; Conservative 0; Mismatches 987; Indels 9; Gaps 1;
QY 530 TCAGGATTTCTTGAAGAGTCCGTGTAGCGAGCGGAGGTAAAGTCTGAGGTTGAGA 589
Db 380 TCGGCAAGCTGCTGCTGCCAGGCGGTGTCGCCAAGGCGCTGGAGAAATCCCGTGGCA 439
QY 590 AGCTTCGTGGGAAAGAGGAAAGTGTAGAGTGTCTTCAAGGGGACACAAATTTTCAAG 649
Db 440 ACTTGGTGGCGGAGAGGCTGAAACGCCGACGTCGAGAGTCCGCGCAGCGGTGG 499
QY 650 CTTTAAAGACTTATGGAAGAGATTGTTGTGAGCAAGCGGAAAGCTTGTATCTGTGATTG 709
Db 500 ACAAGTACACCGTGCACATGACCAAGCGCGCGAGGAAAGCAAGCTCGACCCGCTGATCG 559
QY 710 GTCTGTGATGAGGATTTAGAGAGTCTGTGAGGATTTCTTTCAGGAGACGAGAGCAATC 769
Db 560 GTGCGGACGACGAGATCCGCGGACCATCCAGGTCCTCGACGCGCGGACCAAGAAACACC 619
QY 770 CTGTGCTTTATTTGAGAGCCAGGAGTGTGTAAACAGCTGTGTTTGAAGGTTTAGCAGAAA 829
Db 620 CGTGTCTGATCGGCGAACCCTCGGCGTTCGCGAAGACCGGCATCTGTGAGGGCTTGCCGACG 679
QY 830 GGATTGTGAAGAGAGTGTGCCCAACAGTCTTACTGATGTGAGATTAAATTTGTTGAGCA 889
Db 680 GCATCATCAACGCGGAAAGTGTCCGAGCGGCTCTCAAGGACAAAGCGCTCTGTGCCCCCTGACA 739
QY 890 TGGGTGCGTTAGTTCTGTGCTTAAATACCGAGGAGAGTGTGAAAGAGGTTGAAATCTG 949
Db 740 TGGGGGCGCTGATCCCGGCTGCTCAAGTTCCGCGCGAGTTCGAGGAAACGCTGAAAGCGG 799
QY 950 TTTTGAAGAAAGTTTGAAGACGCTCAAGGCAAGTGTCTCTTTTATTTGATGAGATTCAAT 1009
Db 800 TCCTCAAGAACTGGCGGACGAGGAGGAGGCGGGTCTCTCTTCATCTGACGAACTGACA 859
QY 1010 TGGTTCTTTGTTGGGCAAACTGAAAGGTCGATGAGATGAGCTTAATCTGTTCAAGCCCA 1069
Db 860 CCATGTGCGGCGCGGCAAGCGGAAAGTGTGCTGAGACGCGCGCAACATGCTCAAGCGCG 919
QY 1070 TGTTAGCTAGAGGCGAGCTTCGATGCTTGGTGTCTACAAAGCTTGAAGATACAGGAAT 1129
Db 920 CTCTGGCGCGCGGAGCTGACATGCTGCTGCTGCTACTACCTCGAGAGTATGCGCAGT 979
QY 1130 ATGTTGAGAAAGATGTCGCTTTTGAAGAGGAGTTTCAACAAAGTCTATGTTTGGGAGCAA 1189
Db 980 ACATCGAGAGAGATGCGCGCTGAGCGCGCTTCCAGAAAGTGTGCTGAGCAACCGA 1039
QY 1190 GTGTGCTGACACCATTTAGTATCTTCTAGAGACTCAAGGAGAGATATGAGGAGCATCATG 1249
Db 1040 GCGAGGAGAGACCATCCCATCTCTCGTGGCTCTCAAGGAAACGCTATGAAAGTGCACCAAG 1099
QY 1250 GTGTGCGAATCCAGAGACAGAGCTCTTATAAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1309
Db 1100 GGGTGAAGATCACCGACGCGCGGATCATCCCGCGGCAAGCTGTGCGACCGCTATCATCA 1159
QY 1310 CTGCTCGGCAATTTACCGGATTAAGCAATTTGATTTGTTGATGAGGCTTGTGCGAAATGTA 1369
Db 1160 CCGATCGGCAATGCGCGGACAAAGCCCATCGACTGATCGACGAGGCGCGCGGCGCATCC 1219
QY 1370 GAGTCCAGCTTGTAGTCAACCTTGAGAGATTGATTAACCTTTGAAAGAGAGAGATGAGC 1429
Db 1220 GCATGGAGATCGACTTCAAGCGGAGGAACTGGATGCTCTCGACCGCTGCTGATCCAGC 1279
QY 1430 TGGAAATTTGAACCTTCAAGCTTGAAGAGGAGAGGATTAAGCCAGCAAGCTTCAAGCTTGA 1489
Db 1280 TGAAGATCGAGCGGAGCGGCTGAAGAGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 1339
QY 1490 TAGAGGTGCGGAAAGAGCTTTGATGACCTGAGAGAGAGCTTTCAGGCTCTCAGCAAT 1549

1340 CCAAGCTGGAGGAGATATCGTCAAGCTCGAGCGCAATACGCGACCTCGAGGAGATCT 1399
 1550 ACAGAAAGAGAGAGAGATTTGATGAGATTCGAAGCTTTAAACAGAAAGAGAGAGC 1609
 1400 GGAAGTCGAG 1459
 1610 TCATGTTTCTTTGAG 1669
 1460 CCAAGCAGGAGATGAG 1519
 1670 GATATGCGCGCAATCAAG 1729
 1520 AGTACAG 1579
 1730 AAG 1789
 1580 CCGAG 1639
 1790 GTTGACAG 1849
 1640 AGTGAG 1699
 1850 TTGCTGATAG 1909
 1700 TGGAGCAG 1759
 1910 AGGCAATTTCAAGGTCAG 1969
 1760 AGCGGTGGCGGTTGCGCGCGCGCGCTCGCGATCCGAAACCGCGCGAGCGCTGCTTCC 1819
 1970 TATCTTTGACCAACTGCTGTGTCGCAAACTGAGCTCGCAAGGCTCTTGTGAGCAGC 2029
 1820 TCTCTCTCGCGCGAG 1879
 2030 TGTGATGATGAG 2089
 1880 TCTTCGATACAG 1939
 2090 CTGCTCTCTCGCTCATTTGGGACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2149
 1940 CGGTGGCGCGCTGATCGCGCGCGCTCGCGCTACGTCGGCTTCAGAGAGAGAGAGAGAGAG 1999
 2150 TAACTGAGCTGTGAG 2209
 2000 TGACGAG 2059
 2210 CTCATGTTGCTCTTCAACTCTGCTCAAGTCTTCTCAAGTCTTCTCAAGTCTTCTCAAGTCT 2269
 2060 CCAATCCGAGATGATTTCAACATTTCTCTCAAGTCTTCTCAAGTCTTCTCAAGTCTTCTCAAG 2119
 2270 GGCAG 2329
 2120 GTACAG 2179
 2330 CTGAACACCTCTTGTGAG 2389
 2180 CGGCGAGATCCAGG-----AGTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2239
 2390 TGATCGGAG 2449
 2231 TGATGAG 2289
 2450 TGTGTTTCAG 2509
 2291 TGTGTTTCAG 2349
 2510 AGTGTGCTGTGCGGCTGTGAG 2569
 2351 GCTTGGCAAGGCGCTTGGCGGAG 2410
 2570 ACTATATCTTGGCAG 2629

Db 2411 ACAAGCTGATTGCGCTCGGCTTCGACCCCGTCTATGCGCAGCCCGCTGAAAGCGGCGCA 2470
 Qy 2630 TGCAG 2683
 Db 2471 TCAGCGCTGATTCGAGAAACCCGCTGCGCAATGATCTGCGCGCAAAATTCG 2524

RESULT 10
 US-09-199-637A-272/c
 ; Sequence 272, Application US/09199637A
 ; Patent No. 6355411
 ; GENERAL INFORMATION:
 ; APPLICANT: Ausubel, Frederick
 ; APPLICANT: Goodman, Howard M.
 ; APPLICANT: Rahme, Laurence G.
 ; APPLICANT: Mahajan-Miklos, Shalina
 ; APPLICANT: Tan, Man-Wah
 ; APPLICANT: Cao, Hui
 ; APPLICANT: Drenkard, Eliana
 ; APPLICANT: Tsongalis, John
 ; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
 ; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
 ; FILE REFERENCE: 00786/361002
 ; CURRENT APPLICATION NUMBER: US/09/199, 637A
 ; CURRENT FILING DATE: 1998-11-25
 ; PRIOR APPLICATION NUMBER: 60/066, 517
 ; PRIOR FILING DATE: 1997-11-25
 ; NUMBER OF SEQ ID NOS: 437
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 272
 ; LENGTH: 2970
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-199-637A-272

Query Match 17.6%; Score 546.8; DB 4; Length 2970;
 Best Local Similarity 53.8%; Pred. No. 4.8e-148;
 Matches 1158; Conservative 0; Mismatches 987; Indels 9; Gaps 1;

Qy 530 TCAGGAGATTGTTGAACGAGTGGTGTAGCGAGCGCGAGGGTAAAGTCTGAGGTTGACA 589
 Db 2585 TCAGGAGATTGTTGCTGCTGCGCGAGGGGCTGTCGCGAAGCGCTGAGAGATTCGCGTGC 2526
 Qy 590 AGCTTCGTGGGAAAGAGGAG 649
 Db 2525 ACCTGCTGCGCGAGAGCGGTGAACGACCGAAGCTCGAGGAGTCCGCCAGCGCTGG 2466
 Qy 650 CTTTAAGACTTATGAGAGAGATTTGTTGAGCAGCAGCAGGAGAGTTCGATCTGATG 709
 Db 2465 ACAAGTACACCGTTCGACATGACCAAGCGCGCGAGGAGGCAAGCTCGACCGGTGATCG 2406
 Qy 710 GTGCTGATGAGAGAGATTAGAGAGTCTGAGAGATTTCTTCGAGGAGAAACGAGAAACAATC 769
 Db 2405 GTGCGAGCAGAGATTCGCGCGAGCATCCAGGTCTGCGAGCGCGGAGCAAGAGACACC 2346
 Qy 770 CTGTGCTTATTCGAG 829
 Db 2345 CGGTGCTGATTCGCGGAAACCGCGCTCGGCAAGACCGCATCTGTCGAGGCGCTGGCGCCAGC 2286
 Qy 830 GATTTGAGAGAGAGATGTCGCAACAGCTTACTGATGAGATTAATTTCTGTTGAGCA 889
 Db 2285 GCATCATCAACCGCGAGTCCCGGAGCGGCTCAAGCAAGCGCTGCTGGCGCTTGGACA 2226
 Qy 890 TGGGTGCTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 949
 Db 2225 TGGGGCGCTGATCGCGGTGCAAGTTCGCGCGAGTTCGAGGAGACGCTGAGAGCGG 2166
 Qy 950 TTTTGAAGAGAGTTCGAG 1009
 Db 2165 TCCTCAACGAACCTGGGCAAGCAG 2106
 Qy 1010 TGGTCTTGTGCTGCGCAAAACCTGAAGGAGTGCATGAGTGCAGCTAATCTGTTCAAGCCCA 1069

2105 CCATGTCGGCGCGGCAAGCGAAGGTGCCATGAGCCGCGCAACATGCTCAAGCCGG 2046
1070 TGTAGCTAGAGGCGAGCTTCGATGCTATGCTGCTCAACGCTTGAAGAATACAGGAAT 1129
2045 CTCTGGCGCGCGGAGCTGCACTGCTGCTGCTGCTACTACCTCGACGATATCGCCAGT 1986
1130 ATGTTGAGAAAGATGCTGCTGCTTTCGAGGAGGTTTCCAAAGTCTATGTTGCGGAGCAA 1189
1985 ACATCGAAGAGATGCGCGCTGAGCGCGCTTCCAGAAGGTGCTGCTGAGCAACCGA 1926
1190 GTGTGCTGACACCATAGTATGCTTACGAGCTCAGGAGAGATATGAGGACATCATG 1249
1925 GCGAGAGAGACCATGCTGCTTCCGCTGAGGAGCGCTTCCAGAAGGTGCTGCTGAGCAACCGA 1866
1250 GTGTGCGAATCCAAAGACAGAGCTCTTAAATGCTGCTCAGCTGCTGCTCGTTACATAA 1309
1865 GGTGAGCATCAACGAGCGCGATCATCGCGCGGCAAGCTGTCCACGCTACATCA 1806
1310 CTGTGCGCATTTACCGGATTAAGCAATGATGTTGTTGATGAGGCTTGTGCAATGTGA 1369
1805 CCGATCGGCAACTCGCGGCAAGGCCATCGACCTGATCGACCGCTGCTGATCCAGC 1746
1370 GAGTCCAGCTGTATGATCAACCTGAAAGAGATGTAACCTTGAAGCAAGATGCGAGC 1429
1745 GCATGAGATCGATCTCAAGCCGAGGAACTGGATGCTCTCGACCGCTGCTGATCCAGC 1686
1430 TGGAAATTTGAACTTACGCTTGAAGAGGAGAGGATAAAGCCAGCAAGCTCGACTTA 1489
1685 TGAAGATCGAGCGCGGCGCTGAAGAGGAGAGACGACGAGCCACCAAGAGCGGCTGG 1626
1490 TAGAGTTCGGAAGAGCTTGATGACCTGAGAGACAAAGCTTACGCTTCCAGATGAAT 1549
1625 CCAAGCTGGAAGAGGATGATGCTCAAGCTCGAGCGCGAATACGCGGCACTCGAGGATCT 1566
1550 ACAGAAAGGAGAAAGAGAGATGATGATGAGATTCGAGCTTGAAGAGCTGCTGATTA 1669
1565 GGAAGTCGAGAGAGCGAGGTGAGGCTCGCGGCGAGATCCAGCAGAGATCGAGCAGG 1506
1610 TCATGTTTTCTTTCGAGGAGCGAGAAAGATATGACCTTGAAGAGCTGCTGATTA 1669
1505 CCAAGAGGAGATGAGGCGCGCGCGGCGAGGCGGACCTCGAGAGCATGCGCGCATCC 1446
1670 GATATGCGCAATTCAGAGTGGATCTGCAATGTCGCAATGTCGCAATGTCGCAATGTCG 1729
1445 AGTACAGACCATCCCGGACCTTGAAGCGAGCTGCAAGTGTGCGACGAGCGCAAGA 1386
1730 AAGAAATGATGCTCACAGAAACCTTGGGCTGCAACATGCTGAGGTTGAGGCTGAGCC 1789
1385 CCGAGAACAGTGTGCTGCGCAACAGGTGACCGAGGAAATCGCGGAGTGGTTCCA 1326
1790 GTTGGACAGGATTCAGTCAAGAGCTTGGCCAAATGAGAGGAGAGGTTGATTTGTC 1849
1325 AGTGGACCGGTATCCCGGTGCGAGATGCTCGAGGCGAGCGGAGAGCTGCTCGCA 1266
1850 TTGCTGATGATGATGAGGAGTGTGAGGAGATCAAGCGGTAAATGAGTTTCTG 1909
1265 TGGAGCAGGAGTGTGATCGCGGAGTGTGATCGGCGAGGAGGAGTGTGCGGCTGCA 1206
1910 AGGCAATTTAAGGTCAAGGCGAGGAGTGTGAGGCGCAACAGGCAACTGGATCATCT 1969
1205 ACGCGTGGCGGCTTGGCGCGCGGCTGCGGATCGGACCGCGGCGGCTGCTTCC 1146
1970 TATTCCTGGAACAACTGTTGTTGGCAAACTGAGCTTCGCGCAAGGCTTGTGAGCAGC 2029
1145 TCTTCTCGGCGGACCGGCGTGGCAAGACCGAGTTGTGCAAGGCGTGGCGGAGTTCC 1086
2030 TGTTCGATGATGAACCTTTAGTTGCGATGATGATGTCGGAATATGGAACAACT 2089
1085 TCTTCGATACCGAGGAGGCGCTGTTGCGGATAGATATGTCGAGTTTCATGAGAACT 1026
2090 CTGCTCTCGCTCATTTGGGCGCACACAGGATGTTGTCACAGGAGGAGTGGACAC 2149
1025 CGTGGCGGCGCTGATCGGCGGCGCTCGCGGCTACGTCGGCTCGAGGAGGCGGCTACC 966

QY 2150 TAACTGAGCTGTGAGGAGCGGACCTTATTTGTTGTCATATCTTTGATGAGTGGAGAGG 2209
Db 965 TGACCGAGCGGATCGCGCGCAAGCCCTACTCGGTGGTGTCTGTCGACGAGGTGGAGAGG 906
QY 2210 CTCATGTTCTCTTCAACACTCTGCTCCAAAGTTTGGATGATGATGATGATGACAGAG 2269
Db 905 CCCATCCGATGATTAATCAATCTCTCCAGAGTGTGAGGACGCGCTGACCGACA 846
QY 2270 GGCAGGACGAGACGTCGATTTTCAGAACTCGGTGATTAATCATGACATCAAACTTTGGT 2329
Db 845 GTACGCGGCTACGCTGGAATTTCCGCAACACCGTGGTGTGATGATGATGATGATGATG 786
QY 2330 CTGAACACCTCTCTGCGGCTAACTGCGAAGTAACTGGAAGTGGCCCGGAGTGTG 2389
Db 785 CGCGCGAGATCAGG-----AGCTGGCGCGGACCGCGAGGCGCAACGTGCGCGAG 735
QY 2390 TGATGCGGAGTGAAGAAACACTTTCAGACAGAGACTTTCGAAACAGGCTTTCAGGATG 2449
Db 734 TGATGAGCGGCTCAATGCGCACTTCCGTCGGAATTCATCAACCGGATCGACGAGTGG 675
QY 2450 TGGTGTTCGACCCCTTTCATGACACGAGTGGAGAAAGTAGCTCGGCTTCAAATGAAG 2509
Db 674 TGGTGTTCGAGCGGCTGCTGCGAGCAGATCGCGGATCGCGAGATCCAGCTCGGTC 615
QY 2510 AGTGTGCTGCTCGGCTTGTGAAAGAGGAGTGTGCTTTCGCGAGTCACTGATGCTGCTTGG 2569
Db 614 GCTGCGCAGGCGCTTTCGCGAGCGGAGCTGAGCTGGAATTCAGCCAGGAGGCGCTGG 555
QY 2570 ACTATATCTTGCAGAGAGTATGACCGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 2629
Db 554 ACAAGCTGATTCGCTGCTGCGACCGGCTTATGCGGCGACCGCGCTGAAGCGGCGCA 495
QY 2630 TGGAGAGAGTGTGTCAGAACTGTCAAAGATGTTGTCGTTGAGGAGAAATCG 2683
Db 494 TCCAGCGTGGATGAGAAACCGCTGCGGCAACTGATCTGCGCGGCAAAATTCG 441

RESULT 11

US-09-199-637A-1/c
; Sequence 1, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 42235
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(42235)
; OTHER INFORMATION: N is any nucleic acid.
US-09-199-637A-1

Query Match 17.6%; Score 546.8; DB 4; Length 42235;
Best Local Similarity 53.8%; Pred. No. 2.2e-147;
Matches 1158; Conservative 0; Mismatches 987; Indels 9; Gaps 1;

530 TCAGGATTTGTTGAACGAAGTCGGTGTAGCGACGGCGAGGGTAAAGTCTGAGGTTGAGA 589
25812 TCGGCAAGCTGCTGCTCGCCAGGGGGTGTGCGCGAAGCGCTCGAGAAATGCGTGGCCA 25753
590 AGCTTCGTGGGAAAGAGGAGAAAGTTGAGAGTGTTCAGGGAGACAAATTTTCAAG 649
25752 ACGTCGTGGCGCGAGGCGGTGAACACCGCAACGTCGAGAGTTCGCGCCAGGCGCTGG 25693
650 CTTTAAAGACTTATGGAAGAGATTTGTTGAGCAAGCAGGGAAGCTTGATCTCTGTGATTG 709
25692 ACNAGTACACCGTCGACATGACCAAGCGCGCGAGGAGGCAAGCTCGACCCGCTGATCG 25633
710 GTGCTGATCAGGAGATTAGAAGAGTCTGAGGATTTCTTTGAGGAGAGACGAGAAACAATC 769
25632 GTCGCGACGACGAGATCCCGCCGACCATCCAGGTCTCTGAGCGCGGACCAAGAACAAACC 25573
770 CTGTGCTTATTTGAGAGCCAGAGTTGGTAAACAGCTGTGTTGAAGTTTGAACAAA 829
25572 CGTGTGTCGATCGGCGAACCCGCGTGGCAAGACCGCCATCGTCGAGGGCTTGGCCACG 25513
830 GGATTTGAAAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAAATTTCTTTGGACA 889
25512 GCATCATCAACGCGAAGTTCGCGAGCGCTCAAGGACAAAGCGCTGTGCGCTTGGACA 25453
890 TGGGTGCGTTAGTTGCTGTGTTAAATACCGAGGAGGTTTGAAGAAAGTTGAAATCTG 949
25452 TGGGGGCGCTGATCGCGGTGCCAAGTTCCGCGCGGAGTTTCGAGGAACGCTGAGGCGG 25393
950 TTTTGAAGAAGTTGAGGACGCTGAGGCAAGATGATCTCTTTATTTGATGATGATTCATT 1009
25392 TCTTCAACGAATGGGCAAGCAGGAGGCGGGTCACTCGTGTTCATCGACGAATCGACA 25333
1010 TGGTCTTGGTCTGCGCAAACTGAAGGTCGATGATGACACTAATCTGTTCAGGCCCA 1069
25332 CCAATGTCGCGCGCGCAAGCGGAGGTCATGAGCGCGCAACATGCTCAAGCCGG 25273
1070 TGTAGCTAGAGGCGAGCTTCGATGCAATTTGGTGTCTCAACGCTTGAAGAAATACAGGAAT 1129
25272 CTCTGCGCGCGCGAGCTGCACTCGTGTGTGTACTACCTCGACGAGTATCGCCAGT 25213
1130 ATGTTTGAGAAAGATCTGCTTTGAGAGGAGTTCCAAACAGTCTATGTTGCGGAGCAA 1189
25212 ACATCGAGAAAGATCGCGCTGAGAGCGCGCTTCAGAGGTGCTGTGAGCAACCGA 25153
1190 GTGTGCTGACACCAATTAGTATCTTTAGAGGACTCAAGGAGGATATGAGGACATCATG 1249
25152 GCGAGAGACACCAATCGGCATCTCTCGTGGCTCAAGGAACGCTATGAAGTGACACACG 25093
1250 GTGTGCGAATCCAGACAGACTTTATTAATGCTGCTCAGCTGTCTGCTCGTTACATAA 1309
25092 GGGTGAGCATCACCGACGCGGATCGCGCGCGGCAAGCTGTGCAACCGCTACATCA 25033
1310 CTGGTTCGGCATTTACCGGATAAGCAATTTGATTTGTTGATGAGGCTTGTGCGAATGTA 1369
25032 CCGATCGGCACTCGCGACAGGCCATCGACTGATCGACGAGCGCGCCAGCCGCTCC 24973
1370 GAGTCCAGCTTGATAGTCAACTGAGAGATTGATAACTTTGAAGGAGGAGGATGACG 1429
24972 GCATGAGATCGACTCAAGCGGAGGAACTGGATCGTCTCGACCGTCTGATCCAGC 24913
1430 TGGAAATTTGAATTCAGCCCTTGGAAAGGAGGAGGATTAAGCCAGCAAGGCTCGACTTA 1489
24912 TGAAGATCGAGCGGAGGCGCTGAAGAGGAGAGACGAGAACCCACAGGAAGCGCCCTGG 24953
1490 TAGAGTTCGGAAAGAGCTTGATGACTGAGAGACAAGCTTCAGGCTCTCAGATGAAT 1549
24852 CCAAGCTGAGGAGGATATCGTCAAGCTCGAGCGGATATACGCCACCTTCAGGAGATCT 24793
1550 ACAGAAAGAGAGAGAAATTTGATGATTCGAAGGCTTAAACAGAAAGAGAGAGC 1609
24792 GGAAGTCCGAGAGGCGGAGGTGCGAGGCTCGCGCAGATCCAGCAGAGATCCGAGCAGG 24733

QY 1610 TCATGTTTTCTTTTCAGAGGCGAGAACGAAGATATGATACCTTTGCAAGAGCTGTGATCTAA 1669
Db 24732 CCAAGCAGGAGATGGAGCGCGCGCGCAAGCGCACTTCGAGAGCATGGCGCATCC 24673
QY 1670 GATATGCGCAATTTCAAGAGTGGATCTCAATTTGCCAACTTGAAGNACTTCTTCTG 1729
Db 24672 AGTACCAGACCATCCCGGACCTGGAAACGACGACCTGAGATGGTTCGACACGCAAGA 24613
QY 1730 AAGAGAAATGTGATGCTCACAGAAAACTTTGGGCTCTGAACATTTGCTGAGGTTGTGAGCC 1789
Db 24612 CCGAGAACCAAGTTGCTCGGCAACAAGGTGACCGACGAGGAATTCGCGAGTGTTCCTCA 24553
QY 1790 GTTGGACAGGATTTCCAGTGAACGAGACTTTGGCCAAATGAGAGGAGAGTTGATGGTC 1849
Db 24552 AGTGGACCGGTATCCCGGTGTGGAAGATGTCGAGGCGAGCGGAGAGCTCTGCGCA 24493
QY 1850 TTGCTGATAGTTGATAGCGGGTGTGGGACAGATCAAGCGGTAATTCAGTTCGTTCTG 1909
Db 24492 TGGAGCAGGAGCTGCAATCGCGGAGTGAATCGGCCAGGACGAGGCGGTAGTCGCGGTCTCA 24433
QY 1910 AGGCAATTTCTAAGGTCAAGGGCAGGACTTTGGTAGGCAACAACAGCCAACTTGGATCAATCT 1969
Db 24432 AGCGGTGCGCGCTTCGCGCGCGCGCTCGCCGATCGAACCGCGCGAGCGCTCGTTCC 24373
QY 1970 TATTCCTTGGACCAACTGTGTGTTGGCAAACTGAGCTCGCCAAAGGCTCTTGTGAGCAGC 2029
Db 24372 TCTTCTCGGCCGACCGGGGTGGGCAAGACCGAGTTGTGCAAGGCGCTGCGCGAGTTCC 24313
QY 2030 TGTTCGATGATGAACCTCTTAGTTTCGGATTCATATGTCGGAATATATGGACACACT 2089
Db 24312 TCTTCGATACCGAGGAGCGCTGTCGCGATATATGTCGAGTTTCATGAGGAACACT 24253
QY 2090 CTGTCTCTCCCTCATTTGGGGCACCAAGGGTATGTTGTCACAGGAGGAGGTGGACAAC 2149
Db 24252 CGGTGGCGCGCTGATCGCGCGCTCGGGCTACGTGCGCTTCGAGGAAGCGCGCTACC 24193
QY 2150 TAACTGAGGCTGTGAGAGGCGACTTATTTGTGTCATCTCTTTGATGAGTGGAGAGG 2209
Db 24192 TGACCGAGGCGATCCCGCGCAAGCCCTACTCTCGTGTGTGTCGAGGAGGTGGAGAGG 24133
QY 2210 CTGATGTTGCTGCTTCAACACTCTGCTCCAGTTTGGATGATGTCGATGACAGAG 2269
Db 24132 CCAATCGGATGATTTCAACATTTCTCTCCAGGTGCTCGAGGACGACGCTGACCGACA 24073
QY 2270 GGCAGGACGAGCAGTCGATTTTCAGAACTCGGTGATAATCATGACATCAAACTTTGGTG 2329
Db 24072 GTCAAGGCGGTACGGTGGACTTCGCGCAACACCGTGGTGTGATGACCTCCAACCTCGGTT 24013
QY 2330 CTGAACACCTCTTCAGGCGTAACTGGGAAAGTAACATGGAAGTGGCCCGGACTGTG 2389
Db 24012 CCGCGCAGATCCAGG-----AGCTGGCGGCGAACCGGAGGCGCAACGTGCGCGAG 23962
QY 2390 TGATCGGAGGAGTGAGGAACACTTCAGACCCAGAGCTCTTTGAACAGGCTTGAACGATTTG 2449
Db 23961 TGATGAGCGGCTCAATGCGCACTTCGCTCGGAATTCATCAACCGATCGACGAGTGG 23902
QY 2450 TGGTGTTCGACCCCTTTTCATGACATGACAGTTGAGGAAGTAGCTCGGCTTCAATGAAG 2509
Db 23901 TGGTGTTCGAGCGGCTGGCTCGCGACAGATCGCCGCAATCGCCGAGATCCAGCTCGGTC 23842
QY 2510 ACCTTGTGTCGCGCTTGTGAAAGGAGGAGTTGCTTTTGGCAGTCACTGATGCTGTTTGG 2569
Db 23841 GCCTGCGCAAGCGGCTTGGCGGAGCTGAGCCTGGAATCTGAGCCAGGAGGCGCTGG 23782
QY 2570 ACTATATCTTGGCAGAGATTATGACCCGCTGTATGGTGTCTAGGCCCTATAAGGAGATGA 2629
Db 23781 ACAAGCTGATTGCGCTCGGCTTCGACCCCGTCTATGGCGCACCGCGCTGAGCGGCCA 23722
QY 2630 TGGAGAGAGAGGTGGTGAACAGAACTCTCAAGATGTTGCTGCGTGAAGAAATCG 2683
Db 23721 TCCAGCGCTGATTCGAGAACCCGCTGCGCAACTGATCTCTGCGCGCAAAATTCG 23668

JLT 12

09-252-991A-4842
Sequence 4842, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 4842

LENGTH: 2580

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

09-252-991A-4842

Query March 17, 64; Score 545.2; DB 4; Length 2580;
Best Local Similarity 53.7%; Pred. No. 1.3e-147;
Matches 1157; Conservative 0; Mismatches 988; Indels 9; Gaps 1;

530 TCAGGGATTTGTTCAACGAAGTCGGTGTAGCGCGCGAGGGTAAAGTCTGAGGTTGAGA 589
380 TCAGGAGTCTCTCGCCAGGCGGTGTGCGCAAGGCGCTGGAGATGCGCTGGCCCA 439
590 AGCTTCGTGGGAAAGGAAGGTTGAGAGTGTCTCAGGGGACACAAATTTTCAAG 649
440 ACCTGCGTGGCGGGAAGCGGTGAACACCGACGTCGAGGAGTCGCGCGAGCGCTGG 499
650 CTTTAAAGACTTATGGAAGAGATTGGTTGAGCAAGCAGGAGAGCTTGATTCCTGTGATG 709
500 ACAAAGTACACCGTCGACATGACCAAGCGCGCGAGGAAGCAAGCTCGATCCGCTGATCG 559
710 GTGCTGATGAGGAGATTAGAAGAGTCGTGAGGATTCCTTCGAGGAGAACGAACAATC 769
560 GTCCGAGCAGCAATCCGCGGACCAATCCAGTTCCTGACGCGCGGACCAAGACAACC 619
770 CTGTGCTTATTTGAGAGCCAGGAGTTGGTAAACAGCTGTGTGAGGTTTGAAGCAAA 829
620 CGGTGCTGATCGCGGAACCGCGGTGCGCAAGACCGCCATCGTGGAGGCTCGGCCAGC 679
830 GGATTTGAAGAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTCTGTGACA 889
680 GCATCATCAACGGCGGAAGTGCAGGACCGGCTCAAGGACAAAGCGCTGCTGGCCCTGACA 739
890 TGGTGGCTTAGTTGCTGGTCTAAATACCGAGGAGAGTTTGAAGAAAGGTTGAAATCTG 949
740 TGGGGGCGCTGATCGCGGTGCGCAAGTTCCGCGCGAGTTGAGGAACGCTGAGGAG 799
950 TTTTGAAGAGTTGAGGACGCTGAAGCGAAAGTGTCTCTTTATTTGATGAGATTCATT 1009
800 TCCTCAACGAACTGGGCAAGCAGGAAGCGCGGTCTATCTGTTTCATCGAGCAACTGACA 859
1010 TGGTCTTGTGCTGGCAAACTGAAGGCTGATGATGATGATGATGATGATGATGATGATG 1069
860 CCATGGTGGTGGCGGAAGGCGGAGGTCATGAGCGCGGCAACATGCTCAAGCGCG 919
1070 TGTAGCTAGAGGCGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1129
920 CTCTGGCGCGCGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 979
1130 ATGTTGAGAAGATGCTGCTTTGAGAGAGGTTTCCAAAGTCTATGTTGCGGAGCCAA 1189
980 ACATCGAGAAGATGCGCGCTGAGGCGCGCTTCCAGAAAGGTCGTGTGAGCAACCGA 1039
1190 GTGTGCTGACACATTTAGTATCTTTAGAGGATCTAAGGAGAGATGATGAGGAGATCATG 1249
1040 GCGAGGAGAGACATGATCGGCTCTCGGTGCTCTCAGGAACGCTATGAGTGCACCG 1099

QY 1250 GTGTGCGAATCCAAAGCAGACAGCTCTTATAATGCTGCTCAGCTGTCTGCTGTTACATAA 1309
DB 1100 GGGTGAGCATCACCGACGCGCGATCATCGCGCGGCAAGCTGTGCGACCGCTACATCA 1159
QY 1310 CTGGTGGCATTTACCGGATAAAGCAATTTGATTTGTTGATGAGGCTTGTGGAATTTGA 1369
DB 1160 CCGATCGCAACTGCGGACGAGCCATCGACCTGATCGAGGCGCCGACCCGATCC 1219
QY 1370 GAGTCCAGCTTGTATAGTCAACTGAAGAGATTGATAA CTTTGAAGAGAGAGATGCGAGC 1429
DB 1220 GCATGAGATCGACTCCAAGCCAGAGGAACTGGATCGTCTCGACCGTGGCTGATCCAGC 1279
QY 1430 TCGAATTTGAATCTTCAAGCTTGGAAAGGGAAGGATAAAGCCAGCAAGCTTCACTTA 1489
DB 1280 TGAAGATCGAGCGCGAGCGCTGAAGAAAGGAAGACGACGAGGCCACAGGAGCGCCCTGG 1339
QY 1490 TAGAGTGGGAAAGAGCTTGTATGACCTTGAGAGACAAGCTTCAGCTCTCTCAGATGAAT 1549
DB 1340 CCAAGCTGGAGGAGATATCGTCAAGCTCGAGCGCAATACGCGACCTCGAGGATATCT 1399
QY 1550 ACAGAAAGGAGAAAGAGAGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1609
DB 1400 GGAAGTCCGAGAAGCCCGAGGTGCAAGGCTCGGCGCAGATCCAGCAGAAAGATCGAGCAGG 1459
QY 1610 TCATGTTTTTCTTTCAGGAGGCGAGAAAGATATGACCTTTCAGGAGGAACTTCTTCTG 1669
DB 1460 CCAAGCAGAGATGAGGCGCGCGCGCAAGGCGAGCTCGAGAGATGCGCGCGCATCC 1519
QY 1670 GATATGGCGCAATCAAGAGTGGAACTTGCATTTGCCCAACTTGAAGGAACTTCTTCTG 1729
DB 1520 AGTACAGACCATCCGAGCTTGGAAACGAGCTGAGATGCTGACGACGAGCAGCAAGA 1579
QY 1730 AAGAGATGATGCTTCAAGAAAGCTTTGGGCTGGAACAATTTGCTGAGGTTTGAGCC 1789
DB 1580 CCGAGAACCAAGTTGCTGCGCAACCAAGGTGACCGAGGAAATCGCGAAGTGGTTTCCA 1639
QY 1790 GTTGCAGAGGATTCAGTGTGAGAGACTTGGCCAAATGAGAAGGAGAGTTGATTTGTC 1849
DB 1640 AGTGAACCGATATCCGTTGTGAAAGATGCTCGAGGCGAGCGGAGAGCTGCTGCCA 1699
QY 1850 TTGCTGATAGGTTGATTAAGCGGGTTTGGGACAGATCAAGCGGTAAATGCAAGTTTCTG 1909
DB 1700 TGGAGCAGAGCTGATCGCGAGTGTATCGGCCAGGACGAGCGGTAGTTCGCGGTGTTCA 1759
QY 1910 AGGCAATTTAAGTCAAGGCGAGACTTTGTTAGGACAAACAGCCAACTGGATCATTTCT 1969
DB 1760 AGCGCTGCGCGTTTCGCGCGCGCGCTCGCCGATCCGAACCGGCGAGCGGCTCGTTCC 1819
QY 1970 TATTCCTTGGACCAACTGCTGTGTCGCAAACTGAGCTGCGCAAGGCTCTTCTGAGCAGC 2029
DB 1820 TCTTCTCGGCGCGACCGGGTGGGCAAGACGAGTTGTGCAAGCGCTGGCGAGTTCC 1879
QY 2030 TGTGATGATGAAGAACTTTAGTTGCGATTTGATATGTCGGAATATATGGAACAACACT 2089
DB 1880 TCTTCGATACCGAGGAGCGCTGTTGCGGATCGATATGTCGAGTTTATGAGAGAACT 1939
QY 2090 CTGCTCTCGCTCATTTGGGSCACACAGGCTATGTTGTCACAGGAGGTGGAAC 2149
DB 1940 CGGTGCGCGCTGATTCGCGCGCTTCGCGGCTACGTCGGCTTTGAGGAAGCGGCTACC 1999
QY 2150 TAACTGAGGCTGTGAGGAGGCGACTTATTCGTGTCATCTTTGATGAAGTGAAGG 2209
DB 2000 TGAACGAGGATCCGCGCGAAGCTTACTCGGTGTTGCTGCTGACGAGTGAAGAG 2059
QY 2210 CTGATGTTGCTGCTTCAACACTCTGCTCCAAAGTTTTGGATGATGTTGATGATGATGATG 2269
DB 2060 CCGATCCGATGATTTCAACATTTCTCTCCAGGTGCTCGAGGACGAGCGCTGACCGACA 2119
QY 2270 GCAAGGAGGAGCAGTCAATTTGAGGAATCTGGTGATATCATGATCAATCAACCTTGGTG 2329
DB 2120 GTACGCGGCGTACGCTGAGCTTCGCGCAACACCGTGGTGTGATGATGATGATGATGATG 2179

2330 CTGAACACCTCTTTCAGGCGCTAACTGGGAAAGTAACAATGGAAGTGGCCCGGAGCTGTG 2389
 2180 CGGCACAGATCCAGG-----AGCTGCGCGCGACCGCGGCGCAACGCTGCGCGAG 2230
 2390 TGATGCGGAGGTGAGGAACACTTCAGACCAGAGCTCTTGAAACAGGCTTGAGGAGATTG 2449
 2231 TGATGAGCGCGGTCAATGACACATTCCTCGTCCGGAATTCATCAACCGGATCGACGAAGTGG 2290
 2450 TGGTGTTCGACCCCTTTCACATGACACAGTTGAGGAAAGTAGCTCGGCTTCAAAATGAAG 2509
 2291 TGGTATTCGAGCGCTGGCTCGGAGAGATCGCCGCGATCGCGAGATCCAGCTCGGTC 2350
 2510 ACTGTGCTCCCGCTTGCTGAAAGAGAGTTCCTTTGGCAGTCACTGATGCTGCTTTGG 2569
 2351 GCCTGGCGAAGCCCTGGCGGAGCGGAGCTGAGCCTGGAACCTGAGCCAGGAGCGCTGG 2410
 2570 ACTATATCTGGCAGAGAGTTATGACCCGCTGTATGCTGTAGGCTTATAGGAGATGA 2629
 2411 ACAAGCTGATCGCGCTGGCTTCGACCCCGCTCATGGCGCACCGCCGCTGAAAGCGGGCA 2470
 2630 TGGAGAAGAGGTGGTGAACAGAACTGTCAAAGATGCTTGGCTGAGGAAATCG 2683
 2471 TCCAGCGCTGATCGAGAACCCGCTGGCGCACTGATCTGGCGGCAATTCG 2524

BUILT 13

09-252-991A-4897/c

Sequence 4897, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIORITY FILING DATE: 1999-02-18

PRIORITY FILING DATE: 1998-02-18

PRIORITY FILING DATE: 1998-02-18

PRIORITY FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 4897

LENGTH: 2970

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

09-252-991A-4897

Query Match

Best Local Similarity 53.7%; Pred. No. 1.4e-147;

Matches 1157; Conservative 0; Mismatches 988; Indels 9; Gaps 1;

530 TCAGGATTTGTTGAACGAAGTCTGTTAGCGACGCGGAGGCTAAAGTCTGAGTTGAGA 589

2585 TCGGCAAGCTGTGCTCGCCAGGCGTGTGCGCAAGCGCTGAGGATGCGGTGGCA 2526

590 AGCTTCGTGGGAAAGAGGGAAGATTGAGAGTCTTCAGGAGACAAATTTTCAAG 649

2525 ACCTGCTGCGCGGAGGCGTGAACGACCCGAACTGAGGAGTCTGCGCCAGGCGCTGG 2466

650 CTTTAAAGACTTATGGAAGAGATTGTTGACNAGCAGGAGCTTGATCTCTGTGATTG 709

2465 ACMAGTACCGTCGACATGACCAAGCGCGCGAGGAAGGCAAGCTCGATCCGGTATCG 2406

710 GTCGTGATGAGGAGATTAGAAGAGTCTGAGGATTTCTTTCGAGGAGAACGAGAAACAATC 769

2405 GTCGCGACGAGAAATCCCGCGACCATCCAGGTCTGCGCGCGCGGACCAAGCAACACC 2346

770 CTGTGCTTATGAGAGCGAGAGTTGGTAAACAGCTGTGTTGAGAGTTTACACAAA 829

2345 CGGTGCTGATCGGCGAACCGCGCTGCGCAAGACCGCCATCTGCGAGGCGCTGGCCAGC 2286

830 GGATTCGTGAAGAGAGATGCGCAACAGCTTCTACTGATGTGAGATTAATTTGTTGGACA 889

2285 GCATCATCAACGCGGAAGTGCAGGCGCTCAAGGACAAGCGCTGCTGCGCCCTGGACA 2226
 890 TGGGTGCGTTAGTTGCTGCTGCTAAATACCGAGAGAGTTTGAAGAAAGTTGAAATCTG 949
 2225 TGGGGGCGCTGATCGCGGTGCTCAAGTTTCGCGCGGAGTTCGAGGAACGCTGAAGCAG 2166
 950 TTTTGAAGAAGTTGAGGACGCTCAAGGCAAGTGAATCTCTTTATTGATGAGATTCAAT 1009
 2165 TCCTCAAGCAAGTGGCGCAAGCAGGAAAGCGCGGTATCTCTTTCATCGACGACTGACA 2106
 1010 TGGTCTTGGTGTGCGCAAACTCAAGGTCGATGATGATGATGATGATGATGATGATGAT 1069
 2105 CCATGCTGCTGCGCGCAAGCGGAGGTCGATGAGCGCGGCAACATGCTCAAGCGG 2046
 1070 TGTGAGCTAGAGGCGAGTTCGATGCAATGCTGCTCAACGCTTGAAGATACAGGAAT 1129
 2045 CTCTGCGCGCGCGGAGTTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1886
 1130 ATGTTGAGAAAGATGCTGCTGCTTTCGAGAGGAGTTTCCAAAGTCTATGTTGCGGAGCAA 1189
 1985 ACATCGAAGAGTTCGCGCGCTGAGCGCGCTTCCAGAAAGTGTCTGTCGACGACCGA 1926
 1190 GTGTGCTGACACCAATAGTATCTTTCGAGAGACTCAAGGAGAGTATGAGGAGCATATG 1249
 1925 GCGAGGAAGACACGATCGCCATCTCTGCTGCTCTCAAGGAACGCTATGAAGTGCACACG 1866
 1250 GTGTGCGAATCCAAAGACAGAGCTCTTATAAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1309
 1865 GGGTGAATCACCGAGCGCGGATCATCGCGCGGCAAGCTGTGCGCCGCTATCATCA 1806
 1310 CTGTGCGCATTTACCGGATAAAGCAATGATTTGTTGATGAGGCTTGTGCGAATGTA 1369
 1805 CCGATCGCAACTCGCGGACAGGCGCATCGACCTGATCGAGGCGCGCGCGCATCC 1746
 1370 GAGTCCAGCTTGTAGTCAACCTGAGAGATTTGATTAACCTTGAAGAGAGAGATGAGC 1429
 1745 GCATGAGATCGACTCCAGGCGGAGAACTGATCGTCTGCGCGCTGCTGCTGCTGCTGCT 1686
 1430 TGGAAATTTGAATTCAGCTTCAGGCTTGGAAAGGAGAGGATTAAGCCAGCAAGCTCGACTTA 1489
 1585 TGAAGATCGAGCGCGGCGCTGAGAGGAGAGAGCGAGGAGGAGGAGGAGGAGGAGGAGG 1626
 1490 TAGAGTTCGCGAAAGAGCTTGTGATGACCTGAGAGCAAGCTTTCAGCTTCTCAGCATGAAAT 1549
 1625 CCAAGCTGAGGAGGATATCGTCAAGCTCGAGCGGCAATACGCGGACCTCGAGGATATCT 1566
 1550 ACAGAAAGGAGAGAGAGAGATTCATGAGATTCGAGGCTTAAACGAGAAAGAGAGAGC 1609
 1565 GGAAGTCCGAGAGGCGGAGGTCGAGGCTCGCGCGAGATCCAGGAGAGATCGAGCAGG 1506
 1610 TCATGTTTCTTTTCAGGAGGCGAGAACGAGATATGACCTTTCAGAGAGCTGCTGATCTAA 1669
 1505 CCAGCAGGAGATGAGGCGCGCGCGGCGCAAGGCGGACCTCGAGAGCATGCGCGCATCC 1446
 1670 GATATGCGCAATTCAGAAAGTGAATCTGCAATGCGGCACTTGAAGGAACTTCTTCTG 1729
 1445 AGTACCAAGACCATCCCGGACCTGGAACCGAGCTGAGATGCTGAGGAGGAGGAGGAGG 1386
 1730 AAGAGATGCTGCTCACAGAAACGTTGGGCTGACACATGCTGCTGAGGTTGTGAGCC 1789
 1385 CCGAGAACCGAGTGTGCTGCGCAACAGGAGTGAACGAGGAGAAATCGCGGAGGTTTCCA 1326
 1790 GTTGAACAGGATTCAGTGAACGAGACTTGGCCAAATGAGAGGAGAGAGGTTGATTTGTC 1849
 1325 AGTGACCGGTATCCCGGTGTCGAGATGCTCGAGGCGGAGCGGAGAGGAGTCTGCGCA 1266
 1850 TTGCTGATGAGGTTGATAGCGGTTGTGGAACAGAAATCAAGCGGTAATGAGGTTTCTG 1909
 1265 TGGAGCAGGAGTGTGATCGGCGAGTGTATCGGCGAGGAGGAGGAGGAGGAGGAGGAGG 1206
 1910 AGCAATTTCAAGGTCAGGCGAGGAGTGTGAGGAGCAACAGGCGCACTGAGTCTTCT 1969
 1205 ACGCCGTGCGCGCTTTCGCGCGCGGCTCGCGATCCGAAACCGGCGGAGCGGCTGCTTCC 1146

```

1970 TATTCTTGGACCAACTGGTGTGGCAAACTGAGCTCGCCAGGCTCTTGTGACAGC 2029
1145 TCTTCTCGGCCGCGCGGGTGGCAAGCCAGTGTGTGCAAGGCGCTGGCGAGTTC 1086
2030 TGTTCATGATGAAACCTCTTAGTTCGGATTGATATGTCGGAATATATGGAACACACT 2089
1085 TCTTCATACCGAGGCGCTGTGGGATCGATATGTCGAGTTCATGGAACACT 1026
2090 CTGCTCTCGCTCATTTGGGGACCAACCAAGGATGTTGGTCAAGGAGGTGGACAC 2149
1025 CGGTGGCGGCTGATCGGCGGCTCCGGGCTACGTCGGCTTTGAGGAAGCGGCTACC 966
2150 TAACTGAGCTGTGAGGCGGACCTTATGTGTATCTTTGATGAGTGGAGAAG 2209
965 TGACCGAGGCGATCCCGCGCAAGCCCTACTCGGTGGTGTGTCGACGAGGTGGAGAAG 906
2210 CTATGTTGTCTCTTCAACACTCTCTCCAGTCTTTGGATGATGTCGATTTGACAGAGC 2269
905 CCATCCGGATGATTCACATCTCTCCAGTGTCTCGAGCGACGCTTGACGACA 846
2270 GGCAAGCGAGGACGTCGATTTAGGAACTCGGTGATATCATGATCAAACTTTGGTG 2329
845 GTCAAGCGGCTACGTTGGACTTCCGCAACACCGTGGTGGTATGATGACCTCCAACTCGGTT 786
2330 CTGAACACCTCTTTCAGGCGCTAACTGGGAAAGTAACTGAATGAGTGGCCCGGACTGTG 2389
785 CGCAGATCCAGG-----AGTGGCGGGGACCGCGAGCGCAACGTGCGCGAG 735
2390 TGATCGGAGGTGAGGAAACACTTCAGACACAGAGCTCTTTGAACAGGCTTGACGAGATTG 2449
734 TGATGACGCGGTCAATGACACTTCCGTCCGGATTTCATCAACGGATCGACGAGTGG 675
2450 TGGTGTGACCCCTTCCATGATGACGATTCAGGAAAGTACGTCGGCTTCAAAATGAAG 2509
674 TGGTATTCAGCGCGTGGTTCGCGAGCAGATCGCGCGCATCGCCGAGATCCAGCTCGGTC 615
2510 AGTTGCTGTCGCGCTTCTGAAAGAGGAGTGTCTTGGCAGTCACTGATGCTGTTGG 2569
614 GCTGGGCAAGCGCTGGCGAGCGGAGCTGAGCCTGGAATGAGCCAGGAGGCGCTGG 555
2570 ACTATATCTTGGCAGAGGATATGACCCCGTGTATGCTGCTAGGCTATAGGAGATGA 2629
554 ACAAGCTGATCGCGTTCGCTTCGACCCCGTCTATGGCGCAGCGCCGCTGAGCGGCA 495
2630 TGGAGAGAGGTGTGACAGAACTGTCAAAGATGTTGTGCGTGAGGAATTCG 2683
494 TCCAGCGCTGGATCGAAGACCGGCTGGCGCAACTGATCTGCGCGCAAAATTCG 441

```

LT 14

8-249-380-1

quence 1, Application US/08249380

tent No. 5827685

GENERAL INFORMATION:

APPLICANT: Lindquist, Susan

TITLE OF INVENTION: Methods and Compositions of Genetic

TITLE OF INVENTION: Stress Response Systems

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

CITY: Houston

STATE: TX

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/249,380

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/710,187
; FILING DATE: 31-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3727 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-249-380-1

```

```

Query Match 16.0%; Score 495.4; DB 1; Length 3727;
Best Local Similarity 52.2%; Pred. No. 4.3e-133;
Matches 1351; Conservative 0; Mismatches 1176; Indels 60; Gaps 9;

QY 155 AATCGAAGATGAATCCAGAGAAATTCACACACAGACAAACGAGACAAATTCGTACAGCTC 214
DB 996 AATATATGAACGACCAACGCAATTTACAGAAAGGCTCTAACGATTTTGACGTTGGCTC 1055
QY 215 ATGAGCTAGCTGTGAATGAGGACATGCTCAATTCCTCTCTTTGCAATTTAGCTGGTCTT 274
DB 1056 AAAAATTTGGCTTCGGATCATCAACATCCAAATTAACACATATACATATCTAGCTGCT 1115
QY 275 TGATCTCTGATCCACCGGTATAT-----TTCTTCAAGCAATCTCTAGTGGCGGTGGCG 328
DB 1116 TCATTGAACGCGCAGAGATGGATCAGTCCCTTACCTACAGAACTCTAATTTGAGAGGGCC 1175
QY 329 AGAAGCAGCTCAATCTGTGAAAGAGTATCAATCAAGCCTTGAAGAAAGCTTCCTTAC 388
DB 1176 GTTACGACTATGATCTTTTCAAGAAAGTGGTAAATAGAAATCTAGTAAAGAAATTCCTCAAC 1235
QY 389 AATCTCTCCACCTGATGATATCCAGGAGTCTAGTCTTATTAAGGTCAATTCGTCGTG 448
DB 1236 AGCACTGACCTTCGGAGATATCTCAAGTATGCTTTGGGGAAGTCTCTTCAAGAGC 1295
QY 449 CTCAGCTGCTCAGAGTCAAGAGTGTACTCATTTGGCTGTTGACCACTGATTAATGAG 508
DB 1296 CTGCTAAGATTCAAAACCAACAGAGGACTCATTTATAGCGCAAGACCATATATTTGTTG 1355
QY 509 GTCTTCTTGAAGATCTCAAAATCAGGATTTGTTGAACGAACTCGGTGAGCGACCGCGA 568
DB 1356 CTCTATTCAATGATTCGTCTATTTCAGCAAAATATTTAAGGAAGCTCAAGTAGATATTGAG 1415
QY 569 GGGTAAAGTCTGAGGTTGAGAACTTCGTTGAACTTCGTTGTTGAACTGAAATTTGACTCT 628
DB 1416 CCAATCAAGCAACGAGCTTGAACCTTCGTTGAACTTATGGAAGAGATTTGGTTGACCAAGC 686
QY 629 CAGGGGACACAAATTTTCAAGCTTTAAAGCTTATGGAAGAGATTTGGTTGACCAAGC 686
DB 1476 ATACGAACACACCTTTGGAATATTTTCAAGTACGCCATTTGATATGATGATGAGCAGCTC 1535
QY 687 -----AGGGAAGCTTGAATTCGTTGATTTGGTGTGATGAGGAGATTTAGAGAGTGTGAGGA 742
DB 1536 GTCAAGGTAACTTGACCCCTGTCTATCGCGCGTGAAGAAAGAAATTAAGAGCACTATTAGAG 1595
QY 743 TTCTTTTCGAGGAGACGAAAGAACTTCCTGCTTATTCGAGAGCCAGGAGTTGGTAAAG 802
DB 1596 TTTTAGCAAGAGAAATTAAGTCCAAACCCATGTTTAAATTTGTTGAGCAGGATTCGGTAAAG 1655
QY 803 CAGCTGTGGTTGAAGGTTTAGCAAAAGGATTTGTAAGAGAGATTTGTAAGAGAGATTTG 862
DB 1656 CCGCTATTATTGAAGGTTGCTCAAGAAATCAATTCAGCATGACGTTCCCATTTCTTAC 1715

```

863 CTGATGTGAGATTAATTTTCGTTGGACATATGGGTGCGTTAGTTCCTGGTCTAAATACCGAG 922

1716 AAGCGCTAAATTCCTTCAGTCTAGATTTTGGCCGCAATAACCCGACGTGCTAAATACAAAG 1775

923 GAGAGTTTGAAGAAGAGTTGCAATTCGTTTGAAGAAGCTTGAGACGCTGAAGCGCAAG 982

1776 GTGATTTCGAAGAAGAATTCGAAGGTGTTTGAAGGAATCGAAGATCAAAGACTCTAA 1835

983 TGAATCTCTTAATTGATGAGATTCATTTGGTTCTTGGTGTGCGCAAACTGAAGGGTCGA 1042

1836 TTGTGTATTCAITGATGAATAATCAATGTTAATGGGTAAATGGTAA-----GG 1883

1043 TGGATGCAGCTAAATCTGTTCAAGCCCATGTAGCTAGAGGGCAGCTTCGATCGATTGGTG 1102

1884 ACGACGTGCTAAACATCTTGAAGCCAGCTTTGTTCAGAGGCCCAATTGAAGTCAATCGGTG 1943

1103 CTACAAGCCTTGAAGAATACAGAAATAATGTTGAGAAAGATCTGCCCTTTCAGAGGAGGT 1162

1944 CCACACCAATAACGAATAATAGATCTATTGTGGAAGAAGATGGTGCCTTTTGAAGAAGAT 2003

1163 TCCACACAGTCTAATGTTGCGGACCCAGGTGCTGCCTGACACCAATTAGTATCCTTTAGAGAC 1222

2004 TCCAGAAATTTGAAGTCGCTGGAACCAAGTGTGAGACAAACAGTGGCCATATTGAGAGGTC 2063

1223 TCAAGGAGAAGTATGAGGACATCATGTGTGCGAATCCAAAGACAGAGCTCTTTATAAATG 1282

2064 TGCACCAAGATATGAATACATCATGTTGAAGGATTCGTGATAGCGCTTAGTCACTG 2123

1283 CTGCTCAGCTGTCTGCTGTTATATAAATCTGCTCGCATTTACCGATTAAGCAATTCGATT 1342

2124 CTGCTCAATTTAGCCAAAGCGTTACTTGCATATAGAAGATTTGCCAGATTTCTGCTTTGGATT 2183

1343 TGGTTGATGAGGCTTTGTCCAAATGTGAGAGTCCAGCTGATAGTCAACCTCAAGACGATTG 1402

2184 TAGTTGATATTCTTTGTGCTGTTGCTCGCGTCCGAAGATTCGAAGCAGAGAATTTG 2243

1403 ATACCTTTGAA-AGGAAGAGGATGACGTGGBAAATTGAATTCACGCTTGGAAAGGGAG 1461

2244 ATTCCAAGGAACAGTCAATTTGCAATTCGATTCAGTAGAGATAAAGCTCTAGAGAGAGAG 2303

1462 AAGGATAAAGCCAGCAAGCTCGACTTATAGAGTCCGGAAG-----AGCTTGATGACC 1516

2304 TAGAGTCCGAGTCCACACTAAAGAAGATTTAAGTTAGCTAGACAGAAGGAGCTTCAT 2363

1517 TGAGAGCAAGCTTTACGCTCTCACGATGAAATACAGAAAGAGAGAAAGAGAGAAATTGATG 1576

2364 TGCAAGAAGAAATTTGGAACCTCTAAGACAACGTTTCAATGAAGAAAAGCATGGCCATGAAG 2423

1577 AGATTGGAAGCTTTAAACAGAAAGAGAGAGCTCATGTTTCTTTCGAGAGGCGCAGAAC 1636

2424 AATTGACAACAGCTTAAAGAAAGAAATTTGATGAACTGGAAAAACAAGCCCTTGTAGCTGAAC 2483

1637 GAAGATATGACTTTGCAAGAGCTGCTGATCTAAGATATGGCGCAATTCAGAAAGTGGAAAT 1696

2484 GTAGATATGATACTCGTACCGCGCTGATTTAAGGTACTTCCCATCCAGATATCAAAA 2543

1697 CTGCNAATGCCCCACTTGAAGAACTTTCTTCGAAGAG-----AATG 1738

2544 AGCAAAATCGAAAAGCTTTGAAGATCAGGTTGCTGGAAGAAGAGACGTCGTGGTCCCAACT 2603

1739 TGATGCTTCACAGAAAAGCTTTGGCCCTGAACATATGCTGAGGTTGTGAGCGGTGGACAG 1798

2604 CCATGATCCAAATATGCTGCTGATTCAGACACCATTTCTGAACAGCTGCAAGATTGACTG 2663

1799 GGATTCAGTGAACAGACTTTGGCCAAAATGAGAGGAGAGGTTGATTTGGTCTGCTGATA 1858

2664 GTATCCCTGTTTAAAGAAAGTTCTCAGAAATCTGAAAATGAAAAATGATTCATATGGAACGTG 2723

1859 GGTTCGATAACCGGGTTTGGGACAGNATCAAGCGGTAAATGCAGTTCTTCGAGGCAATTC 1918

2724 ACTTATCATCTGAAGTCTGGGCCAAATGGAATGCCATATAAGCTGTTTTCCAATGCCGTTA 2783

1919 TAAGGCTCAAGGGCAGGACTTGGTAGGGCAACAACAGCCAACTGGATCATTTCTATTCCTTG 1978

2784	Db	GATTGTCTAGATCAGGTTTGTAGCTTAATCCAAGGCAACCAGC---ATCCTTCTTATTTTGTAG	2840
1979	Qy	GACCAACTGGTGTGGCAAAACTGAGCTCGCCAGAGCTCTTGCTGAGCAGCTGTTTGATG	2038
2841	Db	GTTTGTCCGGTTCCGGTAAACTGAATTGGCTAAAAAAGTTGCTGATTTTGTGTTTAATG	2900
2039	Qy	ATGAAAACCTCTTAGTTCGGATTGATATGTCGGAATATATGGAACAACAACCTGCTCTCTC	2098
2901	Db	ATGAGGACATGATGATCAGGTCGATGTGTTCTGAATTAAGCAGAAGTATGCGGTCTCTA	2960
2099	Qy	GCCTCAATTGGGGCAACCAACAGGGTATGTTGGTCAAGGAAGGTGGACAACCTAACTGAGG	2158
2961	Db	AGTTGTTGGGTACCAACGGCAGGTTATGTCCGGGTACGATGAAGGTGGCTTTTAACTAAC	3020
2159	Qy	CTGTGAGGAGGCGACCTTATGTGTCTATACTCTTTTCATCAAGTGGAGAGGCTCATGTTG	2218
3021	Db	AAGTGCAAATACAACCATACTCCGTTTGTGTTATTCGATGAAGTAGAAAAGGCACATCTCG	3080
2219	Qy	CTGCTTTCAACACTCTGCTCCAGTTTTCGATGATGCTCGATTGTGACAGAGCGGCAAGGCA	2278
3081	Db	ATGTTTGTGACTGTCATGCTACAAATGTTGGATGACGGTAGAAATTACTTCTGTTCAAGGTA	3140
2279	Qy	GGACAGTCGATTTTCAGGAACCTCGGTGATTAATCATGACATCAAAACCTTGCTGCTGAACCC	2338
3141	Db	AGACGATCGACTGTTTCCAAATTGTTGTTGTCATCATGACTTCCAAATCTAGGTGCTGAATTTA	3200
2339	Qy	TGCTTCAGGGCTAACTCGGAAGTAGTAACTGGAAGTGGCCCGGAGCTGTGTGATGTCGGG	2398
3201	Db	TCAATTC---TCAACAAGGATCAAAGATCCAAGAACTCTACCAAGAAATTTGTCATGGGTG	3257
2399	Qy	AGGTGAGGAAACACTTTACACAGAGCTCTTCGAACAGGCTTCACAGATTTGTTGGTTCGG	2458
3258	Db	CTGTTAGGCAACATTTACAGCCAGAAATTTTGAACAGAAATTTCTAGTATAGTCATTTTCA	3317
2459	Qy	ACCCCTTTTCATGATCACCACTGTGAGGAAGTAGTCTCGGCTTCAAATGAAAGACGTTGCTG	2518
3318	Db	ACAAGCTATCTAGAAGACTATTCATAAGATCGTGGATATTCGTTGAAGGAAATTTGAAG	3377
2519	Qy	-----TCGGCTTGCTGAAGAGGAGTTGCTTTGGCAGTCACTGATGCTGCTTTGGACT	2572
3378	Db	AGAGATTCGAGCAAAATGATAAACAATTACAAGTTGAAATTTAACTCAAGAGGCCAAGGACT	3437
2573	Qy	ATATCTTGGCAGAGATTTATGACCCGGTGTATGTTGCTCTAGGCTCTATAGGAGATGATGG	2632
3438	Db	TCTTGGCCAAATATGTTGTTATTCGGATGATATGGGTGCACGTCCACTGAACAGGTTAAATTC	3497
2633	Qy	AGAAGAAGGTGGTGCACAGAACTGTCAAGATGTTGTTGCTGAGGAAATTCGATGAAAACT	2692
3498	Db	AAAAAGAAATTTTGAACAACCTGGCCACTAAGGATCTTAAAGAAATGAAATCAAGGATAGG	3557
2693	Qy	CCACTGT 2699	
3558	Db	AAACTGT 3564	

RESULT 15
 US-09-103-840A-2/c
 ; Sequence 2, Application US/09103840A
 ; Patent No. 6294328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.
 ; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ; TITLE OF INVENTION: TUBERCULOSIS
 ; FILE REFERENCE: 24366-20007.00
 ; CURRENT APPLICATION NUMBER: US/09/103,840A
 ; CURRENT FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO. 2

Search completed: February 12, 2004, 22:31:14
Job time : 236 secs

283 GATCCACCGGTATATTTCCCTCAAGCAATCTCTAGTGCCTGGTGGCGAGAACGAGCTCAA 342
121 GATCCACCGGTATATTTCCCTCAAGCAATCTCTAGTGCCTGGTGGCGAGAACGAGCTCAA 180
343 TCTGCTGAAAGAGTGATCAATCAAGCCTTGAAGAAGCTTCCCTTCAATCTCTCCACCT 402
181 TCTGCTGAAAGAGTGATCAATCAAGCCTTGAAGAAGCTTCCCTTCAATCTCTCCACCT 240
403 GATGATATTCAGCGAGTCTAGTCTTATTAAGGTCATTCGTCGCTCAAGCTCTCAG 462
241 GATGATATTCAGCGAGTCTAGTCTTATTAAGGTCATTCGTCGCTCAAGCTCTCAG 300
463 AGTCACGAGGTGATCTCATTTGGCTGTGTGACCAAGTGAATATGGTCTTCTTCAAGAT 522
301 AGTCACGAGGTGATCTCATTTGGCTGTGTGACCAAGTGAATATGGTCTTCTTCAAGAT 360
523 TCTCAATCAGGGAATTTGTGAACGAAGTCGGTGTAGCGACCGCGAGGGTAAAGTCTGAG 582
361 TCTCAATCAGGGAATTTGTGAACGAAGTCGGTGTAGCGACCGCGAGGGTAAAGTCTGAG 420
583 GTTGAGAGCTTCGTGGGAAAGAGGGAAGAAAGTTGAGAGTGTCTTCAAGGGAACAAT 642
421 GTTGAGAGCTTCGTGGGAAAGAGGGAAGAAAGTTGAGAGTGTCTTCAAGGGAACAAT 480
643 TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGACGACGAGCAGGAGCTTGATCCT 702
481 TTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGACGACGAGCAGGAGCTTGATCCT 540
703 GTGATPGTCTGATGAGAGATTAAGAGAGTCTGAGGATTTCTTTCGAGGAGAACGAAG 762
541 GTGATPGTCTGATGAGAGATTAAGAGAGTCTGAGGATTTCTTTCGAGGAGAACGAAG 600
763 AACATCTGCTGCTTATTCGAGAGCAGAGTGTGTAAACAGCTGTGCTGAGGTTTA 822
601 AACATCTGCTGCTTATTCGAGAGCAGAGTGTGTAAACAGCTGTGCTGAGGTTTA 660
823 GCACAAAGGATTTGAAGAGAGATGTGCCCAAGTCTTACTGATGTGAGATTAATTCG 882
661 GCACAAAGGATTTGAAGAGAGATGTGCCCAAGTCTTACTGATGTGAGATTAATTCG 720
883 TTGGACATGGTTCGCTAGTGTCTGCTGCTAAATACCGAGGAGAGTTCGAGAAAGTTG 942
721 TTGGACATGGTTCGCTAGTGTCTGCTGCTAAATACCGAGGAGAGTTCGAGAAAGTTG 780
943 AAATCTGTTTGAAGAGTTCGAGACGCTGAAGGCAAGTGTCTCTTTATTGATGAG 1002
781 AAATCTGTTTGAAGAGTTCGAGACGCTGAAGGCAAGTGTCTCTTTATTGATGAG 840
1003 ATTCAATTTGGTCTTGGTCTGGCAAACTGAAGGCTGATGGATGACGCTAATCTGTTT 1062
841 ATTCAATTTGGTCTTGGTCTGGCAAACTGAAGGCTGATGGATGACGCTAATCTGTTT 900
1063 AAGCCCATGTTAGCTAGAGGCGAGCTTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCT 1122
901 AAGCCCATGTTAGCTAGAGGCGAGCTTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCT 960
1123 AGGAATATGTTGAGAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1182
961 AGGAATATGTTGAGAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
1183 GAGCCAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1242
1021 GAGCCAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
1243 CATCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1302
1081 CATCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
1303 TACATACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1362
1141 TACATACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200

QY 1363 AATGTGAGAGTCCAGCTTGTATAGTCAACCTGAAAGAGATTGATTAACCTTTGAAAGAGAGG 1422
Db 1201 AATGTGAGAGTCCAGCTTGTATAGTCAACCTGAAAGAGATTGATTAACCTTTGAAAGAGAGG 1260
QY 1423 ATGCAGCTGGAATTTGAATTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 1482
Db 1261 ATGCAGCTGGAATTTGAATTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 1320
QY 1483 CGACTTATAGAGTGTGCGAAGAGCTTGTATGACTGAGAGACAAGCTTCAAGCTTCAAGCTTCAAG 1542
Db 1321 CGACTTATAGAGTGTGCGAAGAGCTTGTATGACTGAGAGACAAGCTTCAAGCTTCAAGCTTCAAG 1380
QY 1543 ATGAAATACAGAAAGAGAGAGAGAGATTTGATGATTTGAAAGCTTAAACAGAAAGA 1602
Db 1381 ATGAAATACAGAAAGAGAGAGAGAGATTTGATGATTTGAAAGCTTAAACAGAAAGA 1440
QY 1603 GAAGAGCTCATGTTTCTTTCGAGGAGGAGAGAGAGATTTGATGATTTGAAAGCTTCAAGAGCTGCT 1662
Db 1441 GAAGAGCTCATGTTTCTTTCGAGGAGGAGAGAGATTTGATGATTTGAAAGCTTCAAGAGCTGCT 1500
QY 1663 GATCTAAGATATGCGCAATTTCAAGAGTGAATCTGCAATTTGCCCAACTTTGAAGAACT 1722
Db 1501 GATCTAAGATATGCGCAATTTCAAGAGTGAATCTGCAATTTGCCCAACTTTGAAGAACT 1560
QY 1723 TCTTCTGAAGAGATGTGATGCTCAAGAAACGTTGGGCTGCAACACATTTGCTGAGGTT 1782
Db 1561 TCTTCTGAAGAGATGTGATGCTCAAGAAACGTTGGGCTGCAACACATTTGCTGAGGTT 1620
QY 1783 GTGAGCGCTTGGACAGGATTTCCAGTGAAGAGACTTTGGCCAAATGAGAGAGAGGTTG 1842
Db 1621 GTGAGCGCTTGGACAGGATTTCCAGTGAAGAGACTTTGGCCAAATGAGAGAGAGGTTG 1680
QY 1843 ATTGGCTTCTGATAGTGTGATAGCGGTTGCTGGGACAGAGATCAAGCGGTAAATGCA 1902
Db 1681 ATTGGCTTCTGATAGTGTGATAGCGGTTGCTGGGACAGAGATCAAGCGGTAAATGCA 1740
QY 1903 GTTTCGAGCAATTTCAAGGTCAGAGGAGAGACTTTGGGCAACCAAGAGAGAGGTTG 1962
Db 1741 GTTTCGAGCAATTTCAAGGTCAGAGGAGAGACTTTGGGCAACCAAGAGAGAGGTTG 1800
QY 1963 TCATTCTTATTCCTTGGACCAACTGGTGTGGCAAACTGAGCTGCCAAGGCTTCTGCT 2022
Db 1801 TCATTCTTATTCCTTGGACCAACTGGTGTGGCAAACTGAGCTGCCAAGGCTTCTGCT 1860
QY 2023 GAGCAGCTGTTGATGATGAAACCTTCTAGTTCGAGTGTGATGATGATGATGATGATGATGAT 2082
Db 1861 GAGCAGCTGTTGATGATGAAACCTTCTAGTTCGAGTGTGATGATGATGATGATGATGATGAT 1920
QY 2083 CAACACTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2142
Db 1921 CAACACTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 1980
QY 2143 GGAACAATACTGAGGCTGTGAGGAGGCGACCTTATTTGTTGCTATCTCTTTGATGAGTGTG 2202
Db 1981 GGAACAATACTGAGGCTGTGAGGAGGCGACCTTATTTGTTGCTATCTCTTTGATGAGTGTG 2040
QY 2203 GAGAGGCTCATGCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2262
Db 2041 GAGAGGCTCATGCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2100
QY 2263 ACAGACGGGCAAGGAGGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2322
Db 2101 ACAGACGGGCAAGGAGGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160
QY 2323 CTTTGGTGTGAAACCTCTCTTTCGAGGCTTAACTGGGAAAGTAACTGGAAGTGTGCCCGG 2382
Db 2161 CTTTGGTGTGAAACCTCTCTTTCGAGGCTTAACTGGGAAAGTAACTGGAAGTGTGCCCGG 2220
QY 2383 GACTGTGTGATGCGGAGGCTGAGGAG 2442
Db 2221 GACTGTGTGATGCGGAGGCTGAGGAG 2280
QY 2443 GAGATTTGTGTTTTCGACCCCTTTTCAATGACAGTTCGAGGAAAGTAGTCTCGGCTTCAA 2502

990	CCGCGACGAGGAGATCGCGCGCTCTGTGCGCATTTCTCTCGCGCGCACCAAGAAATAACCC	104
771	TGTGCTTATTGGAGAGCAGGAGCTTGCTTAAACAGCTGTGCTTTGAAGGTTTATGACACAAG	830
1050	CGTCTCTATCGCGAGCCCGCGCTTGGCAGACGCGCGTCTGTGAGGGCTCTCGCGAGCG	1109
831	GATTGTGAAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAAATTTGTTGAGCAT	890
1110	CATCGTTTCGCGCGCAGTGTGCCCAAGTAACCTCTCGACGTCGCGCTCATCGCGCTGCACAT	1169
891	GGGTGCGTTAGTTGCTGCTTAAATACGAGAGGAGTTTGAAGAAAGTTGAAATCTGT	950
1170	GGCGCGCTCGTTCGCGCGCGCCAAAGTACCGCGCGAGTTTCGAGGCGCGCTCAAGCCGT	1229
951	TTTGAAGAAGTTGAGCAGCTGCAAGCGCAAAGTATTCTTTTATTGATGAGATTCAATTT	1010
1230	GCTCAGAGGAGTGGAGAGCGCGAGGGGGAAGTCAATCTTCTCATCGACGAGATACACCT	1289
1011	GGTTCTTGTGTGGCAAAACTGAAGGGTTCGATGGATGAGCTTAATCTGTTCAAGCCCAT	1070
1290	CGTCTCTGGCGCGGCGAGCAGCGAGGGTTCCATGAGCGCGGCCAACCTGTTCAAGCCAAT	1349
1071	GTTAGCTAGAGGGCAGCTTCGATGTCATTTGGTGTCTCAACGCTTGAAGAAATACAGGAAATA	1130
1350	GCTGGCGAGGGGACAGCTCAGGTGTCATCGCGCGCACCGCTGGAGGAGTACCGCAAGTA	1409
1131	TGTTGAGAAAGATGCTGCTTTGAGAGGAGGTTTCCAAAGTCTATGTTGCGGAGCCAAG	1190
1410	CGTGGAGAGGACGACGCTTCGAGCGCGGTTCCAGCAGGTGTTGCTCGCGAGCGCGAG	1469
1191	TGTGCTCTGACACCATTTAGTATGCTTTAGAGACTCAAGGAGAAAGTATGAGGACATCATGG	1250
1470	CGTGCCCGACACCGTCAGCATTTCTGAGGGCGCTCAAGGAGAAAGTACAGGGGCGCCATGG	1529
1251	TGTGCGAATCCGAAGACAGAGCTCTTATAATGCTGCTCAGCTGCTGTCTGCTTACATAAC	1310
1530	CGTGAGGATCCAGGACCGCGCGCTCGTGCTCGCGCACAGCTCTCGCGAGGTACATCAT	1589
1311	TGCTCGGCATTTACCGGATAAAGCAATTCATTTGGTTGATGAGGCTTTGTCGAATGTGAG	1370
1590	GGGTGCGCACTTCCCTGACAAAGCCATAGACTGTGTGACGAGGCTTCGCCCATGTGAG	1649
1371	AGTCCAGCTTGATAGTCAACCTGAAGAGATTGATAACCTTGAAGAGGAGGATGCAGCT	1430
1650	GGTGCAGCTCGACAGCCGCGAGAGATTCATAACCTGGAGAGGAGAGAATCCAGCT	1709
1431	GGAAATTGAACCTTCAGCGCTTGGAAAGGAGAGAGATTAAGCCAGCAAGGCTTCGACTTAT	1490
1710	TGAGGTTGAGCTCCACGCGCTCGAAGAGGAGAGAGCAAGGCCAGCAAGCCCGGCTGAT	1769
1491	AGAGGTGC-----1498	
1770	TGAGGTGCTGTCTCGAGCACTGAATTTCTCTCAACAGTCTCTTGTCTGATGTTCTT	1829
1499	-----GGAAAGAG1506	
1830	GTTTTGAACTTTTGATATTAACGTCGCTTAACTGTGGCCATTTGTTTCAGTTCAGGAAGGAA	1889
1507	CTTGATGACTGAGAGACAAGCTTCAGCGCTCTCACGATGAATAACAGAAAGGAGAAAGAG	1566
1890	TTGGACGATCTGAGGACCAAGCTGCAGCGCCCTTGACCATGAAGTACAGGAAGGAGAGGAG	1949
1567	AGAAATTGATGAGATTCGAAGGCTTAAACAGAAAGAGAGAGCTCATGTTTTTTTTCGAG	1626
1950	AGAAATTGATGAGATCAGGAGCTTGAGCGCGCGGAGGAGCTTCAGTTTCACTCTGCAG	2009
1627	GAGGCGAAGACGAGATATGACCTTGCAGAGCTGCTGATCTTAAGATATGGCGCAATTCAA	1686
2010	GAGGCCGAGCGCGGATGAGCTTGCGCGGTGTGGCGGACCTCAAGTACGGCGCCCTCCAG	2069
1687	GAAGTGGAATCTGGAATTTGCCCAACTTGAAGGAATCTTCTTCTGAAGAGAAATGTGATGCTC	1746
2070	GAAATCGACGCTGCTATCTTCCAAAGTTGGA---GAGCGAAACAGGGGAGAACTGATGCTC	2126

1747 ACAGAAACGTTGGCCCTGAACACATTCCTG----- 1777
2127 ACCGAAACCGTCGGCCCTGAACAAATTCAGAGGTATGTTATTTCTGTTACCGTCA 2186
1778 ----- 1777
2187 CAAAAATTTTCAGAGCAAGTGCAGCAATTTGCCGATCGTCTCTAGTAGAGTAGTCTGTC 2246
1778 ----- AGTTGTGAGCCGTTGGCAGGGATTCC 1805
2247 AGCGTGTGAATGTTGTTCTGCTCTATGCGAGGTGTTGAGCCGTTGGACGGGTATTC 2306
1806 AGTACGAGACTTGGCCAAATAGAGAGGAGGTTGATTCGTTCTGCTGATAGGTTGCA 1865
2307 AGTGACCGGCTTGGCCAGACGCAAGAGAGGCTGGTTGGCCCTGGCTGACAGGCTTCA 2366
1866 TAAGCGGGTTTGGGACAGATCAAGCGGTAAATGCAAGTTTCTGAGGCAATTTCAAGGTC 1925
2367 CCAGAGGGTGGTGGCCAGACAGAGGCTGTGAGGCGCTCGCAGAGGCGGTGCTGAGGTC 2426
1926 AAGGACAGACTTGTAGGGCACACAGCCAACTGGATCATTCTTATTTCTTTGACCAAC 1985
2427 GAGGCGCGCTTGTGACGCGCACACAGCCCACTGGCTGCTCTCTCTTCTTGGTCCGAC 2486
1986 TGGTGTGGCAAACTGAGCTCGCCAAAGGCTCTTGTGAGCAGCTGTTTGTGATGATGAAA 2045
2487 TGGCGTGGGAAAACCTGAGCTGGCCAAAGGCTAGCCGAAACAGCTGTTTCGACGACGAGAA 2546
2046 CCTCTAGTTGGAATGATGTCGGAATATATGGAACAACTCTGCTCTCGCCTCAT 2105
2547 CCTTCTTGTCCGCATCGCATGTGCGAGTACATGAGCAGCACTCGGTTGCCCGCCTCAT 2606
2106 TGGGCAACACACAGGCTATGTT----- 2127
2607 CGGAGCACCACTGGGTAAGTAGCAGAAAATGCATGCTCTGCTATTTTAACTGGGAA 2666
2128 -----GGTCAAGAG 2136
2667 TTGCGATGAACCTGTTCTGACAGTGGGCGGACCTGTGACAGCTACGTCGGCCATGAA 2726
2137 GAAGGTGACAACTAACTAGGCTGTGAGGCGGACCTTATTTGTGTCATCTTTTGTAT 2196
2727 GAGGTTGGGACGTGACTGACAACTGAGGAGGAGGCGGTACAGCGTGATCTCTTTCGAC 2786
2197 GAAGTGGAGAGGCTCATGTTGCTCTTCAACACTCTGCTCCAAAGTTTGGATGATGTT 2256
2787 GAGTTCGAGAGGCCCATGTGCGCGTGTTCACACCTTGCTCCAGGTCCTCGACGAGGC 2846
2257 CGATTGACAGCGGCAAGGACGAGCACTCGATTTTCAGGAACTCGGTGATATCATGACA 2316
2847 AGGCTGACGGATGGCAAGGACGAGGACGCTGAGACTTCAGGAAACCGGTGATCATCATGACA 2906
2317 TCAACCTTGGTGTGACACTCTCTTCAGGCGCTAACTGGGAAAGTAACTCAATGGAAGTG 2376
2907 TCGAACCTTCGGCGCGGACCTCTCTGCTGGGATGTTGGGGAAGAACTCCATGAAGTTC 2966
2377 GCCCGGAGCTGTGTGATGCGG----- 2398
2967 GTCGCGATCTGTTGATGAGGAGGTATGCATTCGATAACGTTCTGAATCGTGTGAGA 3026
2399 -----AGGTGAGGAA 2408
3027 CATTTGTTCTCTGACGCGTGAATGCTCAATTCGGATCGTTATTTGCTGTGTGACAGGTGAGGAG 3086
2409 ACATTTACAGACAGACTTGTGAACAGGCTTACAGAGATTTGTTGTTGACCCCTTTC 2468
3087 GCATTTTCGCGCTGAGCTGCTGAACCGTCTGACAGATCGTGTATTTTCGATCTCTGTC 3146
2469 ACATGACAGTTGAGGAAAGTACGCTCGGCTTCAATGAAAGACGTTGCTGTCGGGCTTGC 2528
3147 CCACGAGGAGCTGAGGAGGTGCTCGCTTCAGATGAAGATGAGGATGAGGCGGCTGCTTGC 3206

QY 2529 TGAAGAGAGAGTGTCTTTGGCAGTCACTGATGCTGCTTTGGACTATATCTTTGGCAGAGAG 2588
Db 3207 CGAAGGGGCACTGCTCTGCTGTGACCGAGCCGCAITGGACATCACTTGTCTCTCTC 3266
QY 2589 TTATGACCC----- 2597
Db 3267 TTACGATCCGGTATGTGACCATCCATGATTTGATCCATCTGAAATTCGTCGGTGACACCTG 3326
QY 2598 -----GGTGTATGTTGCTAGGC 2614
Db 3327 ATGGTGTGACTCTCTTATCTTTCTTGTGTGCTTCAACAACAGGTGTATGCGCGCGGC 3386
QY 2615 CTATAAGGAGATGAGATGGAGAGAAAGGTGGTGACAGAACTGTCAAGATGTTGCGGTG 2674
Db 3387 CAATCAGGAGGTGATTCGAGAAAGAGGTTGTCACGAGCTGTGCAAGATGCTGTATCCAGG 3446
QY 2675 AGGAATCCATCAAACTCCACTGTTTACATAGATG-----CAGGCGCTGGTGATCTTG 2728
Db 3447 AGGATATCACAGAACTGACCGTCTCATCGCCGCGCCGCGCAAGGACGAGCTGG 3506
QY 2729 TGTACCGGGTAGA---AAGTGGAGGTCTAGTGGACGCTTCAACAGGCAAGAGTCAGATG 2785
Db 3507 TCTACAGGTGGACCGGAGCGGCTCTGTTGAACGCTGAGACGGGGATGAAGTCGACA 3566
QY 2786 TGTGATTCATATTTGTAACGGGCCAAAGAGAGTGTATGACGCTCAGCGGTGAAAGAA 2845
Db 3567 TCTGATCAGGTCCCAACAGCTCCACGAGGAGCGCTCGCAGGCGCTCAAGAAAGA 3626
QY 2846 TGAGGATCAGAGAAATAGAGATGACGATAATGAGGA 2882
Db 3627 TGAGGATCATGAGAGGACGAGGACGCGCATGAGCA 3663

RESULT 4

US-09-070-927A-156
; Sequence 156, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512

	ery Match	21.5%;	Score 668.2;	DB 12;	Length 3011208;
	st Local Similarity	57.8%;	Pred. No. 4.5e-176;		
	tches 1239;	Conservative	0;	Mismatches 888;	Indels 18; Gaps 2;
601	AAAGAAGGGAAGAAAGTTGAGAGTCTCTCAGGGGACACAAATTTTCAGGCTTTAAAGACT	660			
2342667	AGAGGAGGGAAGAAAGTGACTTCTCAAAATGAGAGGAAAACATATGAAGCTTTAAACAAA	2342608			
661	TATGGAAGAGATTTTGGTTTGAGCAAGCA-----GGAGAGCTTGATCTCTGTCGATTGCTCGT	714			
2342607	TATGGACGAGATTTTATGTCGGGAGTTAAGACGGGAACTTTGATCAGTGATTTGGACGC	2342548			
715	GATGAGGAGATTAGAAAGAGTCTGTAGGAGTTCTTTTCGAGGGAACGGAAGACAACTCTGTG	774			
2342547	GACGAGAAATTCGTTAATGTCTATCCGAATTTTATCAGAAACAAACAAAATAATCAGTT	2342488			

902

2341407 CTTAAATAGCCGATGTTCTTCATCAAAAAGTAATGGTCAAGACGACCGGTTGAGTTA 2341348
 1903 GTTCTTCTGAGCAATTTCTAAGGTCAGGCGAGGACTTGGTAGGGCAACAGCACTGGA 1962
 2341347 GTTAGTGAGCAGTATTAGTGTCTGCGGGGATTAAGATCCAAAACGACCAATCGGA 2341288
 1963 TCATTCTTATCTCTGAGCAACTGTGTGTGGCAAACTGAGCTCGCCAAAGGCTCTTGTCT 2022
 2341287 TCTTTTATTTCTTAGGCCCAACCGGTGTGTGTAACGGAAGTACAGCAAGCAATTAGCC 2341228
 2023 GAGCAGCTGTTGATGATGAAACACTTTAGTTCGGATTCATATGTCGGAATATATGGA 2082
 2341227 TATAATATGTTGATTTCTGAAGATCATATGATTCGAATTCATATGATGATGAG 2341168
 2083 CAACACTCTGTCTCTCGCTCATTTGGGGCAACCAAGGAGTATGTTGGTCACGAGAAAGT 2142
 2341167 AAACATTCGATCAAGACTTGTGGGGCTCTCCAGGTTATGTTGGATATGAGAAAGC 2341108
 2143 GGCAACTAACTGAGCTGTGAGGAGCGACCTTATTTGTGTCATATCTCTTTGATGAAGTG 2202
 2341107 GGACAACTTACCGAAGCTGAAGACGCAACCCATATTCGATTGTCTTACTTGTGATGAAATC 2341048
 2203 GAGAAGGCTCATGTTGCTGTCTTCAACACTCTGCTCCAGTTTGGATGATGATGATG 2262
 2341047 GAAAGAGCCATCCGACGATTTTAAATATCTTTTACAAAGTACTTGAAGATGGTCGGAAT 2340988
 2263 ACAGCGGCAAGGAGGAGCAGCTCGATTTTCAGAACTCGGTGATATCATGACATCAAAAC 2322
 2340987 ACGGATTCAGAGGGCGTTAATTCGATTTTAAACACCGTAAATCATTTAGAGCTTAAT 2340928
 2323 CTTGGTGCTGAACACTCTCTGAGGCTTAACTGGGAAGTAACTGGAAGTGGCCCGG 2382
 2340927 ATCGGCTCTAATTTTATCTTGAAGAACCGAAGAGGTGAATCTCGCTGAACTAGAA 2340868
 2383 GACTGTGTCATCGGAGGTGAGGAAACACTTCAGACGAGCTCTTGAACAGGCTTGAC 2442
 2340867 TCAGAGGTAAGCAAAATTTGCAAGTGAATTTAAACAGAAATCTTAAACCGGGTAGAT 2340808
 2443 GAGATTGTGGTTCGACCCCTTTTCAATGACCAAGTTGAGGAAAGTAGCTCGGCTTCAA 2502
 2340807 GATAATTTCTATTTAAACCACTTACACTCGCTGATATTAAGGCAATTTGGAAAAATTA 2340748
 2503 ATGAAGAGCTTGTGTCGGCTTCTGTAAGAGGAGTGTCTTTGGCAGTCACTGATGCT 2562
 2340747 GTAGAGAACTTCAAATTCGCTTCTGATCAAGAAATTAACGATTAATTTTCAGATAAC 2340688
 2563 GCTTTGGACTATCTTGGCAGAGAGTTATGACCGGCTGATGTTGCTAGGCTATAGG 2622
 2340687 GCAAGAGCTTTTATGCAAGAGAGCTTATGACCCAGTATACGTTGCGCGCTCCACTGAAA 2340628
 2623 AGATGGATGGAGAGAGGTTGTGACAGAACTGTCAAAGATGGTTGTGCTGAGGAAATC 2682
 2340627 CGATATATGTCGTCATGTCGAAACGCTCTTTCAGAGAGAAATTTGATCTGTTAAAT 2340568
 2683 GATGAAACTCCACTGTTTACATAGATGACGCGCTGTGTATCT 2727
 2340567 ATGCCGCACTCTTCCGTTGAAATTTGATTTACAAGATAAAGAAATTT 2340523

ULT 7
 10-369-493-43449
 sequence 43449, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 43449
 ; LENGTH: 2586
 ; TYPE: DNA
 ; ORGANISM: No. US20030233675A1loc punctiforme
 US-10-369-493-43449

Query Match 21.3%; Score 662; DB 12; Length 2586;
 Best Local Similarity 56.8%; Pred. No. 2.1e-176;
 Matches 1245; Conservative 0; Mismatches 935; Indels 12; Gaps 1;
 QY 531 CAGGGATTTGTTGACCAAGTCGTTAGCGACGCGGAGGGTAAAGTCTGAGGTTGAGAA 590
 DB 372 CAAAGCTTTATTCGAAGATTCGGTTTAGCAAGGCAAACTAAAAGATATTATTAAACA 431
 QY 591 GCTTCGTGGCAAGAGGAGAAAGTTGAGAGTGTTCAGGGGACACAAATTTTCAAGC 650
 DB 432 AGTTCGGGGAGCCAAAGAGTGACCGACCAAAATCCAGAGGCAAAATACGAGCACTGGA 491
 QY 651 TTTAAGACTTTATGGAAGAGATTTGGTTGAGCAAGCAGGGAAGCTTGATCTCTGTGTTGG 710
 DB 492 AAAATACGGCGTGACCTCACAGAACGCGCGCTAAAGGTCAACTCGATCCAGTGTGG 551
 QY 711 TCGTGATGAGGAGATTAGAGAGTCTGTGAGGATTTCTTCGAGGAGAACGAGAACAAATCC 770
 DB 552 GCGGATGATGAGATTCGCGCACTGTCCAAATTTCTGTCTCGCCGCCCAAGAAATTAATCC 611
 QY 771 TGTGCTTTATTTGAGAGCCAGGAGTTGTGTAACACAGCTGTGTTGTAAGGTTTAGCACAAG 830
 DB 612 TGTGCTAATTTGGTGAACCGGCTGTTGGTAAACTGCTATCGCTGAAGGATTAGCACAGCG 671
 QY 831 GATTGTGAAGAGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTTCGTTGGACAT 890
 DB 672 TATTAATGCAAGTGATGTATCCCACTCCCTCAAAGACCGCAAGCTAATTTCTTTAGATAT 731
 QY 891 GGTGCGTTAGTTGCTGTCTTAAATACCGAGGAGGATTTGAGAAAGGTTGAAATCTGT 950
 DB 732 GGGTGTCTTGTGCGGGGCAAAATTCGCGGGTGAATTTGAAGAACGGCTGAAAGCACT 791
 QY 951 TTTGAAGAGATTTGAGAGCGCTCAAGGCAAGAGTGTCTCTTTATTGATGAGATTCATTT 1010
 DB 792 ATTAAGAGATTTACTGAATCTGGCGGCAATTTGTTTATTATTGATGAATTTCAAC 851
 QY 1011 GGTCTTGTGCTGGCAAACTGAAGGTCGATGAGTGCAGCTAATCTCTTCAAGCCCAT 1070
 DB 852 CGTTGTGCGCTGTGTGCAACCCAGCGCGATGATGCGGGTAACTTGTGTTAAACCCGAT 911
 QY 1071 GTTAGCTAGAGGCGAGCTTCGATGCAATTCGTCATCAACGCTTGAAGATACAGGAATA 1130
 DB 912 GTTGGCGGGGTGATTCGCTGTATTTGGCGGCACTCTAGATGAATACCGCAACA 971
 QY 1131 TGTGAGAAAGATGTGCTCTTTGAGAGGAGGTTTCAACAAAGCTTATGTTGCGAGCCCAAG 1190
 DB 972 TATCGAAAGGATGCGCACTAGAAAGACGCTTCCAGCAGGTTTATGTCGATCAACCTAG 1031
 QY 1191 TGTGCTCTGACACCATTTAGTATCCTTAGAGGACTCAAGGAGAGATGAGGGACATCATGG 1250
 DB 1032 TGTAGAGATAGTATTTTCGATTTTGGCGGGTTGAGAGAACGTTATGAAACCCACCGG 1091
 QY 1251 TGTGCGAATCCCAAGACAGAGCTTTTATAAATCTGCTCAGCTCTGCTCTGCTTACATAC 1310
 DB 1092 GGTAAAAATTTCTGATGATGCTTTGGTCCGCCGCCGTTTGTGAGTTCGATATTTAG 1151
 QY 1311 TGTGCGGCAATTTACCGGATTAAGCAATTTGATTTGTTGATGAGGCTTGTGCGAATGTGAG 1370
 DB 1152 CGATCGCTTCTTACCTGATTAAGCCATTGACTTGTGTAGCAAGCGCGCCGACGATTAAA 1211
 QY 1371 AGTCCAGCTTGATAGTCAACCTGGAAGAGATTGATTAACCTTGAAGAGAGAGATCGACT 1430

1212 AATGAGATCACTTCCAAACACAGAAAGTCTGACGAAATGATCGAAGATTCTGCAATT 1271
 1431 GGAATTTGAACTTCAACGCTTGGAAAGGAGGATGATAAGCCAGCAAGCTCGACTTAT 1490
 1272 GGAATTTGAGAGCTTATCGCTGCAAAAAGAGCGATCGCGCTTCTCGTGAACGCTAGA 1331
 1491 AGAGTGGGGAAGAGCTTATGATGACCTGAGAGACAGCTTCAAGCTCTCAGATGAATA 1550
 1332 AAGACTAGAAAAGAAATGCGGATCTCAAGAGAGAACAAAGAACCCCTAAATATCTCAATG 1391
 1551 CAGAAGGAGAAAGAGAGAAATGATGAGATCGAAGGCTTAAACAGAAAAGAGAGAGCT 1610
 1392 GCAATCTGAAAAGATATCAATGACAAATTCATCCGTTAAAAGAGATGAAACGGT 1451
 1611 CATGTTTCTTTGAGAGGAGGAGAACGAGATATGACCTTGAAGAGCTGCTGATTAAG 1670
 1452 CAATTTAGAGATTGAGCAAGCAGAACGATTAACGACCTTAAACGAGCTCGGAGTTGAA 1511
 1671 ATATGGGCAAT-----TCAAGAGTGAATCTGCAATTCGCAATTCGCCAATCTGAAGG 1718
 1512 ATACGGTAATTAAGTTGATGATGCTGATGCTGCAAGAGAGTAGAAGCTGAATTTGGCAAG 1571
 1719 AACTTCTTCTGAAGAGAAATGATGCTCAGAGAAACGTTGGGCTCGAACACATTTGCTGA 1778
 1572 TGCCCAAGAGAGTGAATACTACTATACGGGAGAGAGTACAGAACTGATTTGCTGA 1631
 1779 GGTGTCAGCGGTGGACAGGATTCAGTGAAGAGACTTGGGCCAAATGAGAGAGAG 1838
 1632 AATTAATTTCTAAATGAGAGAGAAATTCCTCATCAGCAAGCTGCTGGAATCTGAGAAAGAGAA 1691
 1839 GTTGATTTGCTTCTCATAGTTGATGATGAGCGGTGTCGAGACAGATCAAGCGGTAAA 1898
 1592 ACTACTGCATTTAGAGATGAATACGCCACCGTGTGATTTGGACAAAGAGAGAGCTAC 1751
 1899 TGCAGTTTCTGAGCAATTTCTAAGGTCAAGGCGAGGACTTGGTAGGCGCAACAGCGCAAC 1958
 1752 AGCGTAGCGATGCAATTCAGCGATCGCGCGTGGACTGCGCGATCCCAATCGTCCAT 1811
 1959 TGGATCATTTCTTCTTGGACCACTGCTGTTGGCAAACTGAGCTGCGCAAGCTCT 2018
 1812 CGTAGCTTTATTTCTTGGGCTTACGGGTGTTGGGTAAACCGAGTTGGCGAAGCGCT 1871
 2019 TGCTGAGCAGCTGTTGATGATGAACCTTTAGTTGCGATTTGATGATGATGATGAT 2078
 1872 GCGGGCTATGTTGATGAGAGAGATGCGCTGTTGCGAATCGATGATGATGATGAT 1931
 2079 GGAACAACTCTGCTCTCGCTCTATGAGGCGACCAACAGGAGTATGTTGGTCAAGGAG 2138
 1932 GGAGAAACACGCGCTCTCCGTTTAACTCGTGGCGCTCCAGGATATGTTGGTTACAGAA 1991
 2139 AGTGGACAACTAACTGAGGCTGTGAGGAGGCGACCTTATTTGTTGATGATGATGATGAT 2198
 1992 AGCGGACAACTAAACAGAGAGGATTCGCGCGCTCTTACTAGTGAATTTCTTTTGAAGA 2051
 2199 AGTGGAGAGGCTCATGTTGCTCTCTTCAACACTCTGCTCCAGTTTTCGATGATGATGATGAT 2258
 2052 AATCGAGAGACACCTGATGTTTAAATCTCTTCTGCAATTTCTCGATGATGATGATGAT 2111
 2259 ATTGACAGAGGCGCAAGGAGGAGCACTGATTTTCAAGAACTCGGTGATGATGATGATGATGAT 2318
 2112 CGTCACTGATGCTCAAGGTCAATGAGTGGACTTCAGAAATGCTATTTATTTATCATGACTAG 2171
 2319 AAACCTTGGTGTGACACCTCTTTCAGGCGTAACTGGGAAAGTAACTGAAAGTGGC 2378
 2172 CAACATCGGTTCGCAATATATTTCTGATGCTGCGTGGGATTAATGCTCACTAGCAGCAAT 2231
 2379 CCGGACTGTGATGCGGAGGTGAGGAAACACTTTCAGACAGAGCTTTGAAAGGCT 2438
 2232 GCGCGTCACTGATGAGAGCGATGCGAATAGCTTTCGCTCAGAAATTTCTCAACCGGAT 2291
 2439 TGAAGAGATTGTTGGTTTCCAGCCCTTTTCACTGACCAAGTGGAGAAAGTACGCTCGCT 2498
 2292 TGACGAAATCATCTTCCACGGTTTAGATAAGAGGATTTGCGGAGATTGCTGTT 2351

2499 TCAATGAAGAGAGCTTGTCTCGGCTTCTGTAAGAGAGAGTGTCTTGGCAGTCACTGA 2558
 2352 GCAAGTACAAAGATTAGCCAAAGATTGCCGATCGCAAAATATCCCTCAAGCTCTCAGA 2411
 2559 TGTCTGTTGGACTATATCTTGGCAGAGAGTATGACCCGGTGTATGTTGCTAGGCTAT 2618
 2412 TGTGCACTGACTTTTATGAGCAAGTATGATGATGATGATGATGATGATGATGATGAT 2471
 2619 AAGGAGATGGATGGAGAGAGGTTGTTGACAGACTGTCAAGAGTGTGTTGCTGAGGA 2678
 2472 GAAACGGGATTCAGCGAGAGTGAATCAAAATGCAAAAGCATCTTGGCGGTGA 2531
 2679 AATCGATGAAACTCCACTGTTTACATAGT 2710
 2532 ATTCACGATGGCAACACCATCTTTGTAGTG 2563

RESULT 8
 US-10-369-493-42135
 ; Sequence 42135, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 42135
 ; LENGTH: 2604
 ; TYPE: DNA
 ; ORGANISM: Lactococcus lactis
 ; US-10-369-493-42135

Query Match 21.2%; Score 659.8; DB 12; Length 2604;
 Best Local Similarity 58.1%; Pred. No. 8.9e-176;
 Matches 1225; Conservative 0; Mismatches 872; Indels 12; Gaps 3;

604 GAAGGAGAAAGTTGAGAGTCTTCAAGGGAGACAAATTTCAAGCTTTAAAGACTTAT 663
 439 GAGGTGATAAAGTGACCCAGTCAAAATGCAAGAGAAACATACAAAGCACTTGAATAAT 498
 664 GGAAGAGATTTGGTTGAGCAAG-----CAGGGAAGCTTGTATCTGTTGATGTTGCTGAT 717
 499 GGGTAGATCTCGTTGCTCAAGTTAAATCAGTAAACAGATCTCTCATTTGGACGTGAT 558
 718 GAGGAGATTAGAGAGTGTGAGGATTTCTTTCGAGGAGAACGAGAAACAATCTCTGCTT 777
 559 GAAGAAATTCGTGATGTCAATTCGAGTGTCTTCTCTGTAACAAATAACCCGCTTCTT 618
 778 ATTGAGAGCCAGGAGTTGTTGAAACAGCTGTGTTGAAGGTTTAGCACAAGAGTTGTG 837
 619 ATTGTTGAACCTCGGGTGTGTTAAACAGCCATTTGTTGAAGGATTTGGCACAAGAGTTGC 678
 838 AAGGAGATGTGCCCAACAGTCTTACTGATGTGAGATTAATTTTCGTTGGACATGGGTGG 897
 679 AGAAAGACGTTCTGAAATCTGAAAGATAAACAATCTTTTCACTTGATATGGGTGCC 738
 898 TTAGTCTGCTGCTAAATACCGAGGAGAGTTTGAAGAAAGGTTGAAATCTGTTTGA 957
 739 CTGATTCAGAGCGCAATATCTGTTGATTTTGAAGAGCTTTGAAGCAGTCTTAA 798
 958 GAGTTGAGAGCTGAGGCAAGGATTTCTTTTATGATGAGATTTCAATTTGGTCTT 1017
 799 GAGTGAATAAATCAGATGAGCAAAATTTATCTTCTTTTATGATGAACCTTCATACGATTT 858

1018 GGTGCTGGCAAACTGAAAGGTCGATGAGTGAAGCTAAATCTGTTCAAGCCCACTGTTAGCT 1077
 859 GGTGCAAGTAAACAGAGAGTTCAATGAGAGCTGGGAACCTTTTAAACCAATGCTTGA 918
 1078 AGAGGCGAGCTTCGATGCAATGTTGCTCAACGCTTGAAGATACAGGAATATGTTGAG 1137
 919 CGTGGTGAATCTCATTTGATTTGGGCACTACCTTGGATGAATATCGTAAATACATGGA 978
 1138 AAGATGCTGCTTTGAGAGAGGTTTCAACAGTCTATGTTGCGGAGCAAGTGTGCT 1197
 979 ACAGATAAGAGCACTTGAACGCTGTTTCCAAAGGTATTTGTTACTGAACCTACTGTTGA 1038
 1198 GACACATTTAGTATCTTACAGAGATCTCAAGAGAGATATGAGGACATCATGTTGTGGA 1257
 1039 GATACAAATTTCAATCTTGGCTGCTTAAAGAACGTTTGAATATCCACATGAGTGAAC 1098
 1258 ATCCAGAGCAGAGCTTATATAATGCTGCTCAGCTGCTGCTGCTTACATTAATGCTGG 1317
 1099 ATTATGATATGCTTTAGTCGAGCAGCAACCTTTCAATCGTTATATTAATGACCGA 1158
 1318 CATTTACCGATTAAGCAATTAATTTGTTGATGAGGCTTGTGCGAATGTGAGAGTCCAG 1377
 1159 TTTTACAGATAAAGCCATTTGACTTGTGTTGATGAAGCAGTGAACAAATTCGTGTAGA 1218
 1378 CTGTGATGTCACCTGAAGAGATTTGATTAACCTTTGAAGAGAGAGGAGCTGGAATTT 1437
 1219 ATGATTTCACTCCCACTGAACTTGAACCAAGCTAATCGTCTGCTGATGCAATTAAGAA 1278
 1438 GAATTTACGCTTGAAGAGGAGAGGATTAAGCCAGCAAGCTCGACTTATAGAGGTG 1497
 1279 GAAGAGCGGCTTCAAAAGAGAGAGATGATGCTTCTAGAAACGCTTGAATATATA 1338
 1498 CGGAAGAGCTTGTGATGATGAGAGAGAGCTTCAAGCTTCAAGATGAAATACAGAAAG 1557
 1339 CGTGTGAAATTTGCTGAACTTGAAGAGAGAGATTAATCAACTCAAGCTCAATGGAAGCT 1398
 1558 GAGAAGAGAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1617
 1399 GAGAAGAGAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1458
 1618 TCTTTCAGGAGGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1734
 1459 GAATTTGAGAGAGCTCAAAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1518
 1678 GCAATTCAG 1734
 1519 AAAATTCCTGAAATTTGAAG 1578
 1735 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1794
 1579 TTATCCTTAGTTCAAGATTCAGTCACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1638
 1795 ACAGGATTTCCAGTGAAG 1854
 1639 ACAGGATTTCCAAATTTACCAATTTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1698
 1855 GATAGTTGATTAAGCGGTTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1914
 1699 GAACACATGCTCAATCAAGAGGCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1758
 1915 ATTCTAAGTCAAGGAG 1974
 1759 ATTATCTGCGCTGAG 1818
 1975 CTGGAACCACTGTTGTTGCAAACTGAGCTGCGCAAGGCTTCTGCTGAGAGAGAGAGAGAG 2034
 1819 CTGCTGCAACTGCGGAGTGAAG 1878
 2035 GATGATGAAG 2094
 1879 GATTGAG 1938

2095 TCTCGCTCATTTGGGCGACACACAGGCTATGTTGGTCAACGAGAGAGTGGACACACTAACT 2154
 1939 TCTCGTTAGTTCGAGAGACCTTCCAGGATATGTCGGTTACGATGAAGAGGCGCAATTGACT 1998
 2155 GAGCTGTGAGAGGCGGACCTTATTTGTTGTCATATCTCTTTGATGAAGTGGAGAGGCTCAT 2214
 1999 GAAGCTGTGCTGTAATCTTATACAAATCATCTTCTGCTTGAATCGAAGAGACAT 2058
 2215 GTTCTGCTCTTCAACATCTCTGCTCCAAAGTTTGGATGATGATGATGATGATGATGATGAT 2274
 2059 CCAGATGCTCTCAATATCTTGTGCAAGTTTGGATGATGATGATGATGATGATGATGATGAT 2118
 2275 GGCAGGAGAGTTCGATTTTCAAGAACTCGGTGATATTAATCATGATCAACACCTTGTGCTGAA 2334
 2119 GGTGCTCTGTTGATTTTGAAGACACAGTGTGTTGATTTGATTTGATTTGATTTGATTTGAT 2178
 2335 CACCTCTCTGAGGCTTAACTGGGAAAGTAACTAACT ---GGAAGTGGCCCGGAGCTGTGTG 2391
 2179 TATTTGCTGATTAATGTTGAGAGAAATGGTGAATTTCCGAGAGAAACAACTGAGAAAGGTG 2238
 2392 ATGCGGAGGAGTGGAGAACACTTTCAGACACAGAGCTTTCGACAGGCTTGCAGAGATTGTG 2451
 2239 ATGCTCAACTTCCGGGACATTTTCAACCCAGAAATTTTCAATCGAATTTGATGATGATGAT 2298
 2452 GTGTTGACACCCCTTTTCAATGACACAGTTCGAGAAAGTAACTCGGCTTCAATGAGAAAGAC 2511
 2299 CTTTTCAACCACTTGGCACTTGAAGATTTTGAAGATTTATCGTTAAATGATGATGATGATGAT 2358
 2512 GTTGTCTCGGCTTGTGTTGAAGAGAGAGTGTGTTTGGAGTCACTGATGCTGCTTTGGAC 2571
 2359 CTTTTCACTGCTTCTGAAGAAATGAGATTTCACTTGAAGTGAAGAGTAAAGTT 2418
 2572 TATATCTTGCAGAGAGTATGATGACCGGCTGTTGTTGAGTGTAGGCTTATAGGAGATGATG 2631
 2419 TGGATTTCCGAAATGCTTATGAAACAGCTTATGTTGTCAGCTCACTCAACGTTATTTG 2478
 2632 GAGAAGAGTGTGTGACAGAACTGTCAAGAGATGTTTGTGCTGAGGAAATCGATGAGAAAC 2691
 2479 ACAAAGTCACTGAGAAATCCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2538
 2692 TCCACTGTT 2700
 2539 TCAAAAGTT 2547

RESULT 9
 US-10-310-154-117
 ; Sequence 117, Application US/10310154
 ; Publication No. US20030233670A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Edgerton, Michael D
 ; APPLICANT: Chomet, Paul S.
 ; APPLICANT: Adams, Thomas H
 ; APPLICANT: Ruff, Thomas G.
 ; APPLICANT: Agarwal, Ameeta K.
 ; APPLICANT: Ahrens, Jeffrey E.
 ; APPLICANT: Ball, James A.
 ; APPLICANT: Banu, G.
 ; APPLICANT: Bell, Erin
 ; APPLICANT: Boddupalli, Raghava
 ; APPLICANT: Deikman, Jill
 ; APPLICANT: Deng, Molian
 ; APPLICANT: Dong, Jinzhao
 ; APPLICANT: Duff, Stephen M.
 ; APPLICANT: Galligan, Meghan M.
 ; APPLICANT: Hinchey, Brenda S.
 ; APPLICANT: Huang, Shihshieh
 ; APPLICANT: Johnson, G. Richard
 ; APPLICANT: Jung, Vincent
 ; APPLICANT: Kretzmer, Keith A.
 ; APPLICANT: Laccetti, Lucille B.
 ; APPLICANT: Lai, Chao-Qiang
 ; APPLICANT: Lee, Gary

PPLICANT: Lin, Jie-Yi
 PPLICANT: Liu, Jingdong
 PPLICANT: Lu, Bin
 PPLICANT: Luethy, Michael M.
 PPLICANT: Lund, Adrian
 PPLICANT: Madson, Linda L.
 PPLICANT: Malloy, Kathleen A.
 PPLICANT: McKiel, Christine L.
 PPLICANT: Miller, Philip W.
 PPLICANT: Padmavathi, Manchikanti
 PPLICANT: Start, William G.
 PPLICANT: Tennesen, Dan
 PPLICANT: Vidya, K.R.
 PPLICANT: Wang, Haiyun
 PPLICANT: Xin, Zhanqun
 PPLICANT: Xu, Nanfei
 PPLICANT: Yang, Chunzhi
 PPLICANT: Zeng, Xiaoping
 PPLICANT: Zhang, Qiang
 PPLICANT: Zhao, Yajuan
 PPLICANT: Zhou, Li
 TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
 FILE REFERENCE: 38-15(52796)B
 URRENT APPLICATION NUMBER: US/10/310,154
 URRENT FILING DATE: 2002-12-04
 RIOR APPLICATION NUMBER: 60/337,358
 RIOR FILING DATE: 2001-12-04
 UMBER OF SEQ ID NOS: 736
 Q ID NO 117
 LENGTH: 3000
 TYPE: DNA
 ORGANISM: Xylella fastidiosa
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (187)...(2769)
 OTHER INFORMATION:
 0-310-154-117
 396 TCCACCTGATGATATCCAGCAGTCTCAGTCTTATTAAGTCAATCGTCGCTCAAGC 455
 414 TCAGACGGTAAATGCTCACCAGCAATGAATGAGTCGTTGTTTCATCGACTGATAA 473
 456 TGCTCAGAGTCAAGAGTGTACTCAATTTGGCTGTGACCAAGTTGATTATGGGCTTCT 515
 474 GTTAGCGCAGCAGATGGCGATCAGTTTATGCCAGTGAAGTGTGCTGTTGGCTGTGGT 533
 516 TGAAGATTCT---CAATCAGGATTTGTTGAACGAAGTTCGCTGAGCAGCGGAGGGT 572
 534 CGATGATAGTGGGGGGCTGGGCGTGGCTGCTGCTGCTGGTGGGAAAAAAGAGAT 593
 573 AAAGTCTGAGTTGAGAGCTTCGTTGGGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 632
 594 TGAGGCTGGATGATATAATCGTGGCGGTGAAACTGTCCAGACTGAGAAATGCCAGGA 653
 633 GCACACAAATTTTCAAGCTTTAAGACTTTATCGAAGAGATTTGGTTGAGCAGCAGGAA 692
 654 GCAGCGTCAAGGTTGGAAGAGTACAGATTGATCTGACTGCCAGGCTGAGAGTGGCA 713
 693 GCTTGATCTGTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 752
 714 GCTTGATCCGTTGATGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 773
 753 GAGACGAGAGACAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 812
 774 GCGTCAAAAAATAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 833
 813 TGAAGGTTTAGCACAAGAGATTTGAAGAGAGATGTGCCACAGCTCTTACTGATGTGAG 872

Db 834 AGAAGGGCTGGCTCAGCGCATTTGTCAATGGTGGAGTTCCGGAAGGGTTGCCAGCTAAGCG 893
 Qy 873 ATTAATTTTCGTTGGACATGGCTGGCTAGTTGCTGGTCTAAATACCGAGGAGAGTTTGA 932
 Db 894 TCTGCTCTCGCTGGATTTGGAGCGTTGATTTGCTGGGCGCAAGTTTCGTGGTGGATTGA 953
 Qy 933 AGAAAGGTTGAAATCTGTTTGAAGAGTTGAGGACGCTGAAGCGCAAGTGAATTTCTCTT 992
 Db 954 GAGCGCTTGAAGGGGGTCTTAACGATCTCGCTAAGAAATGAGGGGGGGTCAATTTGTT 1013
 Qy 993 TATTGATGAGATTCAATTTGGTCTTGGTCTGGGCAAACTGAAGGGTGCATGATGAGC 1052
 Db 1014 CATTGACGAGCTGCATACCATGTTGGTGGCGGTAAGCCGATGCTGCGATGATGCTGG 1073
 Qy 1053 TAACTGTTCAAGCCCATGTTAGCTAGAGGCGACCTTCGATGCTGCTGCTCAACGCT 1112
 Db 1074 CATTATGCTCAAGCGGGGTTAGCAGTGGTGAATTTGATGTTATCGGTGGACTAGCTT 1133
 Qy 1113 TGAAGAAATACAGGAAATATGTTGAGAAAGATGCTTGCCTTTGAGAGGAGTTTCCAAAGT 1172
 Db 1134 GGATGAGTATCGCAAGTACATTGAGAAAGGATCGCGCTTGAAGCGCGCTTCCAGAAAGT 1193
 Qy 1173 CTATGTTCCGAGCGCAAGTGTGCTGACCAATAGTATCTTTAGAGAGACTCAAGAGAA 1232
 Db 1194 GTTTGTTGGGAGCGCGACGCGTGGAGGATACCATCGCAATTTCTTCGCGACTTAAGGAAA 1253
 Qy 1233 GTATGAGGCGCATCATGCTGTGCGAATCCAAAGACAGAGCTCTTATAAATGCTGCTCAGCT 1292
 Db 1254 GTATGCGTTGCACACCGGTGTGGAATCACTGATCGGCTATTGTTGCTCGGCTAGCTT 1313
 Qy 1293 GTCTGCTGTTACATAAATCTGGTGGCAATTTACCGGATPAAAGCAATTTGTTGTTGATGA 1352
 Db 1314 GTCTAATCGCTACATCATGCTGATCGTCAAGTACCAGATAAAGCGATTGATGATGA 1373
 Qy 1353 GGCTTGTGCGAATGTGAGAGTCCAGCTTGTATGATGATCACTGAGAGAGATTGATTAACCTTGA 1412
 Db 1374 GGCTGCCAGTGTATTCGTATGGAATTTGACTCCAGCGGAGAGAGCTTGTGCTGTTGA 1433
 Qy 1413 AAGAAAGAGGATGAGCTGGAAATTTGAATTTCAAGCTTGGAAAGGAGAGAGATTAAGC 1472
 Db 1434 GCGTGGTGTGTTTCAATTTGAGTGAATTTCAAGCTGAGATGCTGAAGAAAGAGATGAGGC 1493
 Qy 1473 CAGCAAGCTCGACTTATAGAGTGGGAAAGAGCTTGTATGATGAGAGACAAAGCTTCA 1532
 Db 1494 GAGCAAGCAGCTTTAGCGGATCTTGAAGCTGATATTGAGTTCTGGACCGTGAAATTTTC 1553
 Qy 1533 GCCTCTCAAGTGAATACAGAAAGAGAGAGAGAGATTTGATGAGATTGAGAGGCTTAA 1592
 Db 1554 CGATCTGAGAGAGTGTGGAGATCAGAAAGCGCGCTACAGGGGGCGACTAAGATCAA 1613
 Qy 1593 ACAGAAAGAGAGAGAGCTCATGTTTCTTTTTCAGAGGAGCAGAACGAAGATATGACCTTGC 1652
 Db 1614 GGATCGATCGAGCAGCGCAAGCTTGTATTGGAAGCGCGCAGCGCGCTCAGACTACGC 1673
 Qy 1653 AAGAGCTGCTGATCTAAGATATGGCGCAATTTCAAGAGTGGAAATCTGCAATTTGCCAAT 1712
 Db 1674 CAAGATGAGCGAGATTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1733
 Qy 1713 TGAAGGAATCTTCTTGAAGAGAAATGATGCTGCTCAGAAAGAGCTTGGGCGCTGACACAT 1772
 Db 1734 AAGTCAGCGCGAAGCAGCAGATTTTACGTTGGTGGCAGGAGAAAGTACTGCTGAGGAGAT 1793
 Qy 1773 TGCTGAGGTTGTGAGCGGCTTGGACAGGATTTCAAGTGAAGAGACTTGGCGCAAAATGAGAA 1832
 Db 1794 TGCCGAGGTAGTCAAGTGGTGGACTGCTGATTTCCAGTGAAGAGATGCTTGGGGCGAGCG 1853
 Qy 1833 GAGAGGTTGATTTGGTCTTGTGATAGGTTGCAATAAGCGGTTGTGGGACAGAACTCAAGC 1892
 Db 1854 CGATAAGTTCTTGGTATGGAAGCTGATCTAGCGCGACGCTGCTGCTGCTGCTGCTGCTG 1913
 Qy 1893 GGTAAATGCAAGTTCTTGGAGCAATTTCAAGGTCAGGGCAGGACTTGGTAGGGCACAA 1952
 Db 1914 GATCAAGGTTGATCGATGCGGTACCGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1973

1953 GCCAATGATCATTTCTTATTCCTTGGACCAACTGGTGTGGCAAACTGAGCTCGCAA 2012
 1974 ACCGAGCGCTGTTTCTTCTGGTCCGAGCGGTGTGGTAAGACTGATGATGTA 2033
 2013 GCGTCTTGTGAGAGCTGTTGTGATGATGAACCTCTTGTGATGATGATGATG 2072
 2034 GCGCTGCTGATTTTGTTCGACAGTCAAGATGCAATGCTGCGCATGATGATG 2093
 2073 ATATATGGAACAACTCTCTCTCGCTCATTTGGGACACACACAGGATGATG 2132
 2094 GTTCATGGAAGCAATTCGTGGCGGCTGATTTGGTGGCTTCCGGCTATG 2153
 2133 CGAGAAAGTGGACAACTAATGAGGCTGTGAGAGGCGACCTTATTTGTGATCT 2192
 2154 TGAGGAAGGAGGTTTCTTACTGATTTGGTGGCAGCTCGGCTTACTCTGATCT 2213
 2193 TGATGAAGTGGAGAGGCTCATTTGCTGCTTCTCAACTCTGCTCCAGTTTGG 2252
 2214 GGATGAAGTGGAGAGGCGCATAGTGTGTTCAATATTTCTGCTGCAAGTGA 2273
 2253 TGGTCAATGACAGAGCGGCAAGGAGGAGTGTGATTTCAAGAACTCGGTGATA 2312
 2274 TGGAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2333
 2313 GACATCAAACTTGGTGTGATGATGATGATGATGATGATGATGATGATGAT 2372
 2334 GACATCAAACTTGGTGTGATGATGATGATGATGATGATGATGATGATGAT 2393
 2373 AGTGGCCC-----GGGACTGTGATGATGATGATGATGATGATGATGAT 2426
 2394 GTATACGCAATGAGGCGGAGTGTGATGATGATGATGATGATGATGATGAT 2453
 2427 CTGACAGGCTTGACAGAGTGTGATGATGATGATGATGATGATGATGATGAT 2486
 2454 CATTAACAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2513
 2487 AGTACGCTTCAATGAAGAGCTGTGATGATGATGATGATGATGATGATGAT 2546
 2514 GATTGCGGAGTCAACTGACAGGCGCTTGAGAGCGTCTGCTGAATCTGATGAA 2573
 2547 GCGAGTCACTGATGCTGTTGGATATATCTTGGCAGAGATGATGATGATGAT 2606
 2574 AGATTGATGATGCTGCTGTTGAAATGCTTGTGATGATGATGATGATGATG 2633
 2607 TGCTAGGCTTAAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 2666
 2634 TGCCGCGCGCTGAGAGAGCGATCCATCTCAGCTGAGAGATCCCTTAGCGAG 2693
 2667 TGTCGCTGAGGAATCGATGAATACTCCACCTGTTTATGATGATGATGATGAT 2726
 2694 CTTAGCTGAAGCGTTCGTCAGTGGAGATGATGATGATGATGATGATGATGAT 2753
 2727 TGCTACCGGCTAGAAAGTGGAGTCTAGTGA 2759
 2754 TGTAATCGTAAGTGTGAAGTTGTTTACTGTA 2786

JUL 10
 10-369-493-41201
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 41201
 LENGTH: 2586
 TYPE: DNA
 ORGANISM: *Xylella fastidiosa*
 US-10-369-493-41201

Query Match 21.0%; Score 651.4; DB 12; Length 2586;
 Best Local Similarity 55.5%; Pred. No. 2.1e-173;
 Matches 1302; Conservative 0; Mismatches 1036; Indels 9; Gaps 2;

396 TCACCTGATGATATTCACGCGAGTTCTAGTCTTATTAAAGGTCATTTCGTCGTCCTCAACG 455
 228 TCAGACGGTTAATGCTCTACCCAGCAATGAATTGAGTCGTTTGTTCATCCGACTGATAA 287
 456 TGCTCAGAGTCAAGAGGTGATCTATTTGCTGTGTGACAGTTGATGATGATGATGATGAT 515
 288 GTTAGCGCAGCAGCATGGCGATCAGTTTATGCCAGTGAAGTTCGTTGCTGTTGCTGTTG 347
 516 TGAAGATTCT---CAAATCAGGATTTGTTGAACGAGTCGCTGTAGCAGCGCGAGGT 572
 348 CGATGATAGTGGGCGGCTGGCGCAGCGCTGCTGCTGCTGCGGAAAAAAGAAAGAT 407
 573 AAAGTCTGAGGTTGAGAGCTTCGTGGGAAAAAGAGGAAAGTTGAGAGTGTCTTCAGG 632
 408 TGAAGCTGCGATGATTAATTCGTGGCGGTGAATCTGTCAGACTGAGATGCGGAGA 467
 633 GACACAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTTGTTGAGCAAGCAGGAA 692
 468 GCAGCGCTCAAGGTTTGGAAAAAGTACACGATTTGATCTGACTGCCAGGCTGAGAGTGC 527
 693 GCTTCACTGCTGATTTGCTGCTGATGAGAGATTAAGAGAGTCTGTCAGAGTTCCTTCGAG 752
 528 GCTTCACTGCTGATTTGCTGCTGATGAGAGATTAAGAGAGTCTGTCAGAGTTCCTTCGAG 587
 753 GAGACGAGAAACAATCTCTGCTTATTTGAGAGCAGAGTTGTTAAACAGCTGTCTGT 812
 588 GGTACAAAAAATATCTGCTGCTGATTTGTTGAGCGCGGTGTTGTTGTTGTTGTTGTTG 647
 813 TGAAGTTTGAACAAAGATTTGTTGAGAGAGATGTTGCCAACAGTCTTATCTGATGTTGAG 872
 648 AGAAGGCTGCTCAGCGCATTTGTTCAATGTTGAGGTTCCGGAAGGTTTCCGAGTAAGCG 707
 873 ATTAATTTGTTGAGATGCTGCTGCTGATGAGAGATTAAGAGAGTCTGTTGTTGTTGTTG 932
 708 TCTGCTCTGCTGATTTGGGAGCGCTTCAATGCTGGGCGCAAGTTTCGTTGTTGTTGTTG 767
 933 AGAAAGTTGAAATCTGTTTGAAGAGATTTGAGAGCGCTGAAAGCAAGTGAATCTCTTT 992
 768 GAGCGCTTGAAGCGGCTGTTAAAGATCTGCTAAAGATGAGGCGCGGTCAATTTGTT 827
 993 TATTGATGAGATTCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1052
 828 CATTACGAGCTGCAATACCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 887
 1053 TAATCTGTTCAAGCCCATGTTAGCTGAGAGCGAGCTTCGATGCTGATGCTGCTGCTGCTG 1112
 888 CAATATGCTCAAGCGCGGTGATGACAGTGTGTTGATTTGTTGTTGTTGTTGTTGTTGTT 947
 1113 TGAAGAAATACAGGAATAATGTTGAGAAAGATGCTGCTTTCGTTTTCGAGAGAGGTTTCC 1172
 948 GGATGAGTATCGAAGTACATTTGAGAGGATGCGCGCTTGGAACGCGTTTCCAGAGAGT 1007
 1173 CTATGTTGGGAGCCAGTGTGCTGACACCATTAAGTATCTTATGAGAGATCTCAAGAGAA 1232
 1008 GTTTGTTGGGAGCCAGGTTGAGAGATACCATGCTCAATTTCTGTTGAGTAAAGAGAAA 1057
 1233 GTATGAGGAGCATCATGTTGCGAAATCCAGAGCAGAGCTCTTATAAATGCTGCTGAGCT 1292
 1068 GTATGCGTTGACCCAGCGTGTGGAATACTGATCCGCGCTATTTGTTGCTGCGCTACGCT 1127

1293 GTCTGCTGTTTACATATCTGCTGCGCATTTTACCGGATAAAGCAATGATTGTTGATGA 1352
1128 GTCTTAATTCGCTACATCACTGATCGTCACTAGTTACCAGATAAAGCGATTGACTTGTGATGA 1187
1353 GGCTTGTGCGAATCTGAGAGTCCAGAGCTTGTATGATCAACCTGAGAGAGATTGATCACTTGA 1412
11188 GGCTGCGAGTCTGATTCGTTAGGAATTTGACTCCAGCCGGAAGAGCTTGTATCGTTTGA 1247
1413 AAGGAAGAGGATGAGCTGCGAATTTGAACCTTCACGCTTGGAAAGGAGAGATAAAGC 1472
1248 GCCTCGCTTCAATTCAGTTGAAATTCAGCGTGAGATGCTGAAGAAAGAAAGATGAGC 1307
1473 CAGCAAGCTCGACTTATAGAGTGGGAAAGAGCTTGTATGAGCTCAGAGAGACAAGCTTCA 1532
1308 GAGCAAGCAGCGTTTACCGCATCTTGAGCGGTGATTTGAGGTTCTGCAACCGTGAATTTTC 1367
1533 GCCTCTCACGATGAATACAGAAAGAGAGAGAGAAATTTGATGAGATTCGAGAGCTTAA 1592
1368 CGATCTGGAGAGGTGGAGATCAGAAAGCCGCTACAGGGGGCGACTAAGATCAA 1427
1593 ACAGAAAGAGAGAGCTCATGTTTCTTTGAGGAGGAGAGAGAGAGAGAGATATGACCTTGC 1652
1428 GGAGTCTGATCAGAGAGCGCAAGCTTGTATTTGGAAGCCGCGCAGCGCTCAGACTACGC 1487
1653 AAGAGCTGCTGATCTAAGATATGCGCAATTCAGAAAGTGGAAATCTGCAATTCGCCCACT 1712
1488 CAAGATGAGCGAGATTCAGTATGCTGTACTCCCGCGTTTGGAGAGCAGCTGTGGCAGC 1547
1713 TGAAGAACTTCTTGAAGAGAAATGTGATGCTCACAGAAAGCTTGGGCTGAAACAT 1772
1548 AAGTCAGGCGGACAGACGATTTTACGTTGTTGTCAGGAGAAAGTCACTGCTGAGAGAT 1607
1773 TGCTGAGTTGTGAGCGGTTGAGAGAGATTCAGTGACGAGACTTTGGCCAAATAGAA 1832
1608 TGCCGAGGTAGTCACTCGTTGGACTGGTATTCAGTGAGCAAGATCTTTGAGGGGAGCG 1667
1833 GAGAGGTTGATTTGCTGCTGATAGTTGATAGCGGTTGTCAGGAGAGATCAAGATCAAG 1892
1668 CGATAAGTTGCTGATGAGAGCTGATCTAGCGGAGCTGTGGTTGTTGTTCAAGAGAGGC 1727
1893 GGTAAATGCAATTTCTGAGCAATTTCTAGGTCACAGGAGGAGGACTTTGTAGGCGCAACA 1952
1728 GATCAAGTGTGATCGAGTGGGTAGCGGTGCGGTTCCGCTACTGTTGCTGTGATCCAAATCG 1787
1953 GCAACTGAGATCACTTTATTCCTTGGACCAACTGTTGTTGGCAAACTGAGCTCGCCAA 2012
1788 ACCGAGCGGCTCGTTCCTTTCTGGTCCGACGGGTGTTGTAAGACTGAGTTATGTA 1847
2013 GGCTCTGCTGAGCAGCTGTTGATGATGAACCTTTAGTTGCGATTTGATGTCGGA 2072
1848 GCGCTGGCTGAATTTTGTGACAGTCAAGATGCAATGGTCCGCAATTTGATGATGATGA 1907
2073 ATATATGAACAACACTCTGCTCTCGCCTCAATTTGGGCGACACAGGATATTTGGTCA 2132
1908 GTTATGGAAGAAAGCACTTGTGGCGGCTGATTTGTTGGCTCCGGCTATGTTGGTTA 1967
2133 CGAGGAAGGTGAGCAACTTAATTCCTTGGACCAACTGTTGTTGGCAAACTGAGCTCGCCAA 2192
1968 TGAGGAAGGAGGTTATCTTACTGAAATTTGGTGGACGCTCGGCTTACTCTCTGATCTTT 2027
2193 TGATGAAGTGGAGAGGCTGATGTTGCTCTTCAACACTCTGCTCCAAAGTTTGGATGA 2252
2028 GGAAGTGGAGAGGCGGATAGTATGTTGTTCAATATTTCTGCTGCAAGTACTTGTGA 2087
2253 TGCTCGATTTGACAGCGGCAAGGAGGAGGAGTGGATTTTCAGGAAGCTGGTGTATATCAT 2312
2088 TGAGCGCTTGACTGATGTTGTCAGGCGGCTGATGTTGATTTTCCATATACCGTTATTTGAT 2147
2313 GACATCAAACTTTGGTGTGAACACTCTTCCAGGGCTTAATGGGAAAGTAACTGAGTCTTCA 2372
2148 GACATCGAACTGGGTTCACATCAGATCCAGAACTCAGTGGAGAGCGATTTCTCCGGAGT 2207
2373 AGTGGGCC-----GGGACTGTGTGATGCGGAGGTGAGGAAACACTTCAGACACAGAGCT 2426

Db 2208 GTATACGCAAAATGAAGCGCGCAGTGCATGAGGTGTGAGGGCGCATTTTCGCCACAGATT 2267
QY 2427 CTTGAAACAGGCTTGAAGAGATTGTTGTTTCAACCCCTTCAATGACACCAAGTTCAGGAA 2486
Db 2268 CATTAAACAGGTTGGATGATATTTGTTTTCACCCGTTGGACAGGCGCAATCAAGCA 2327
QY 2487 AGTAGCTCGGCTTCAAAATGAAGAGCTTGTCTCGGCTTGTGCTGAAAGGAGTGTCTTT 2546
Db 2328 GATTGCGGGATTCACTGCGGGCTTGAAGAGCTCTCGCTGAATCTGAGTTGAAACT 2387
QY 2547 GGCAGTCACTGATGCTGCTTTGGACTATATCTTGCAGAGAGTTATGACCCGCTGTATGG 2606
Db 2388 AGATTGATGATGATGCTGCTTGGAAATCTGTTGTTAACTCGGAATTTGATCTGTATATGG 2447
QY 2607 TGCTAGGCTTATAGGAGATGATGAGAGAGAGAGTGTGACAGAACTGTCAAGAGTGT 2666
Db 2448 TGCCCGCCGCTGAGAGAGAGAGATCCAACTCTCAGCTGGAGAAATCCCTTAGCGCAGAGAT 2507
QY 2667 TGTGCTGAGGAAATTCGATGAAACTCCACCTGTTTACATAGATCAGGCGCTGGTGATCT 2726
Db 2508 CTTAGCTGAGGCTTCTGCTGAGAGATCTGTGAGGTGGTGTGATGAGGAGGAAAGCT 2567
QY 2727 TGTGTAC 2733
Db 2568 TGTATTTC 2574

RESULT 11
US-10-369-493-32853
; Sequence 32853, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 32853
; LENGTH: 2571
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-369-493-32853

Query Match 21.0%; Score 650.6; DB 12; Length 2571;
Best Local Similarity 55.5%; Pred. No. 3.6e-173;
Matches 1300; Conservative 0; Mismatches 1034; Indels 9; Gaps 2;

QY 396 TCCACCTGATGATATTCACGAGATTCTAGTCTTATTATAGGTCACTGCTGCTCAAGC 455
Db 228 TCAGACGGTTAATGCTCACCCAGCAATGAATGAGTCGTTTTCATCGGACTGATAA 287
QY 456 TGCTCAGAGTCAAGGATGATCACTATTGGCTGTTGACCACTGATTGATGGTCTTCT 515
Db 288 GTTAGCGCAGCAGCATGGTGCATCAGTTTATGGCCAGTGAAGTGTGTTGCTGTGCT 347
QY 516 TGAAGATTCT---CAATCAGGATTTGTTGAACGAGTGGTGTGACGCGGAGGCT 572
Db 348 CGATGATAGTGGGGGGCTGGGCGGCGTTGCGTGTGCTGCTGCGGAAAAAAGAGAT 407
QY 573 AAGTCTGAGTGTGAGAGCTCTGCTGGGAAAGAGGAGAGAGATTGAGAGTCTTCAG 632
Db 408 TGAGGCTGCGATTGATGAATTTGCTGGCGGTGAACTGTCAGACTGAGATCGGAGGA 467
QY 633 GGACACAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTTGTTGTCAGCAAGCAGGAA 692

468 GCAGCGTCAAGCGTTGAAAGAGTACACGATTGATCTGACTCCAGGCGCTGAGAGTGCAG 527
593 GCTTGATCTCTGATGATTGGTCTGATGAGGAGATAGAGAGTCTGAGGATTTCTTCGAG 752
528 GCTTGATCCGCGTATGATGATGATGAGGAAATTCGCGCAGCAGCAGATATTCGAGCG 587
753 GAGAACGAGAAACAATCTCTGCTTTATTTGAGAGCCAGGAGTTGGTAAACACAGCTGGT 812
588 GCCTACAAAATATATCTCTGCTGATTTGGTGAGCCGCGTGTGGTAAACCTGCCATTGT 647
813 TGAAGGTTTACGACAAAGGATTTGAAAGGAGATGTGCCCAACAGTCTTACTGATGTAG 872
648 AGAAGGCGTGGCTCAGCGCATTTGCAATGGTGGAGTTCCGGAAGGGTTGCGCAGTAAGCG 707
873 ATTAATTTCTGTTGACATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 932
708 TCTGCTCTCGCTGATTTGGGGCGTTGATTTGGGGCGCAAGTTTCGTGGTGGATTCGA 767
933 AGAAGGTTGAAATCTGTTTGAAGAGATTTGAGGACGCTGAAGGCAAGTGTATCTCTT 992
768 GGAGCGCTTGAAGGGGTGCTTAACGATCTCGCTAAGATGAGGGCGGCTCATTTGTT 827
993 TATTGATGAGATTCATTTGGTCTTGGTGTGGCAAACTGAAGGGTGCATGTGATGAGC 1052
828 CATTTGATGAGCTGCATACCATGTTGGTGGCGGTAAAGCCGATGGTGGATGATGCTGG 887
1053 TAATCTGTTCAAGCCCATGTTAGCTAGAGGCGAGCTTCGATGCAATGGTCTACACGCT 1112
888 CAATATGCTCAAGCCGCGTTAGCAGCTGGTGGATTCGATTTGATCGGTGGCATAGTT 947
1113 TGAAGAAATACAGGAAATATGTTGAAGAGATGCTGCTTTGAGAGGAGGTTTCAACAAGT 1172
948 GGATGATATCGCAAGTACATTTGAGAGGATGCGGCTTGGAGCGCGTTTCCAGAGGT 1007
1173 CTATGTTGCGAGCAAGTGTGCTGACACCATTTAGTATCCTTTAGAGGATTCAGAGGAA 1232
1008 GTTTGTTGGGAGCGGAGCGTGGAGGATACCATCGCAATTTCTTGGTGGATTAAGGAA 1067
1233 GTATGAGGACATCATGTTGGCAATTCAGAGAGAGCTCTTATTAATGCTGCTCAGCT 1292
1068 GTATGCTGTCACACCGTGTGGAATACATGATCCGCTATTTGCTCGGCTACGCT 1127
1293 GTCTGCTCGTTACATACTGTCGCGCAATTTACCGGATAAGCAATTTGATTTGATGA 1352
1128 GTCTAATCGCTACATCACTGATGCTGATTTACAGATAAAGCGATTCGATTTGATGA 1187
1353 GGCTGTGCGNATGTGAGAGTCCAGCTGTGATGATCACTGAGAGATTTGATTAACCTTGA 1412
1188 GGCTGCCAGTCTGATTCGTATGGAATTTGACTCCAAACCGGAAGAGCTTGTATGTTGA 1247
1413 AAGGAAGAGGATGAGCTGGAAATTTGAACTTTCAGCGCTTGGAAAGGAGAGGATTAAGC 1472
1248 GCGTCCGCTGATTCAGTTGAAATTTACGCTGAGATGCTGAAAGAAAGAGATGAGGC 1307
1473 CAGCAAGCTCGATTTAGAGTGTGCGAAAGAGCTGTGATGATCTGAGAGCAAGCTTCA 1532
1308 GAGCAAGCAGCGTTTACTGCTATCTGAGCGTGTATTTGAGGTTCTGAGCCGCTGAATTTTC 1367
1533 GCCTCTCAGNATGAATACAGAAAGGAGAAAGAGAGATTTGATGAGATTCAGAGCTTAA 1592
1368 CGATCTGGAGAGGTGTGGAGATCAGAAAAAGCGCGCTACAGGGGCGGCTAAGATCAA 1427
1593 ACAGAAAAGAGAGAGCTCATGTTTCTTTGAGGAGCAGAACGAGAGATGACCTTGC 1652
1428 GGATGCTGATCAGCAGGCCAAGTGTGATTTGGAGCGCGCGCAGCGCTCAGAGCTACGC 1487
1653 AAGAGCTGCTGATCTAAGATATGCGCAATTCAGAAAGTGGAAATCTGCAATTCGCCAACT 1712
1488 CAAGATGAGCGAGATTCAGATGTTGTTACTCCCGCTTGGAGAGCAGCTAGTGGCAGC 1547
1713 TGAAGAACTTCTTCTGAGAGATGTGATGCTCAGAAAAAGTGGGCGCTGAACAT 1772

1548 AAGTCAGCGGAAACAGCAGATTTTACTTTTGGTCAGGAGAAAGTGACTGCTGAGGAGAT 1607
1773 TGCTGAGGTTTGTAGCGCTTGGACAGGGATTTCCAGTGAAGAGCTTGGCCCAAAATGAGAA 1832
1608 TGCCGAGGTAGTCACTCGTTGGACTGGTATTTCCAGTGAGCAAGATGCTTGGGGCGAGCG 1667
1833 CGAGAGGTTGATTTGCTTGTCTGATAGGTGTCATAGCGGGTGTGGGACAGAAATCAAGC 1892
1668 CGATTAAGTTGTTGCGTATGGAAGCTGATCTAGGTGCGCTGCTGGTTCAGATGAGGC 1727
1893 GGTAAATGAGCTTTCTGAGGCAATTTAAGGTCAAGGCGAGAGCTTGGTAGGGCAACA 1952
1728 TATCAAGGTGATCGATCGGTGCGGCTTCCGCTACTGCTGCTGATCCAAATCG 1787
1953 GCCAACTGATCATTTCTTATTTCTTGGACCAACTGGTGTGGCAAACTGAGCTCGCCAA 2012
1788 ACCGAGCGCTCGTTCTCTTTCTGGGTCGACGCGGTGTTGGTAAGACTGAGTTATGTAA 1847
2013 GGCTCTTGTGAGCAGCTGTTTGTATGATGAATAAACCTTCTTAGTTGCGATTTGATGTCGA 2072
1848 GCGCTGCTGATTTTGTTCGACAGTCAAGTCAANTGGTCCGATTTGATAGTGA 1907
2073 ATATATGAAACAACACTCTGCTCTGCTCTGCTGCTCACTGCTCCTCAAGTTTGGATGA 2132
1908 GTTCAATGAAAAACAATTTCTGCGCGCTGATTTGGTGCACCTCCGGCTATGTGGTTA 1967
2133 CGAGGAAGGTGACAACTAACTGAGGCTGTCAGGAGGCGACTTATTTGCTGCTACTCTT 2192
1968 TGAGGAGAGGTTATCTTCTGAAATGGCGGAGCTCGGCTTACTCTCTGATTTCTTT 2027
2193 TGATGAAGTGAAGAGGCTCATGTTGCTCTTCAACACTCTGCTCCTCAAGTTTGGATGA 2252
2028 GGATGAAGTGAAGAGGCGCATAGTGTGTTCAATATTTCTGCTGCAAGTACTTGA 2087
2253 TGCTGATGACAGACGGGCAAGGAGGACAGTGCATTTGAGGAACCTCGGTGATTAATCAT 2312
2088 TGAGCGCTGATGATGCTGAGGCGCTGATGTTGATTTTCCGCAATACCGTTATTTGAT 2147
2313 GACATCAAACTTGTGCTGAACACTCTCTGCGGGCTAACTGGGAAGTAAACAATGGA 2372
2148 GACATCAAACTTAGGTTTACATCAGATCCAAAGACTCAGTGGAGACGATTTCTCGGAGT 2207
2373 AGTGGCCC-----GGGACTGTGTGATCGGAGGTGAGGAAACAATTCAGACCAGAGCT 2426
2208 GTATACGCAAAATGAAGCGCGCATGATGGGGTGGTGCAGGTGCAATTTTCGCCAGAAAT 2267
2427 CTTGACAGGCTTGTGAGAGATTTGCTGTTGACCCCTTTTCAATGACAGCTGAGGAA 2486
2268 CATTAAAGGTTGGATGATTTGTTTTCACCCGTTGGAACAGGCGCAAAATCAAGCA 2327
2487 AGTAGCTCGGCTTCAAAATGAAGAGCTTGTCTCGGCTTGTGAAAGAGGATTTGCTTT 2546
2328 GATTCGCGGATTTCACTCGGGGGCTAGAGAGGCTTCTGCTGATCTGATGTTGAACT 2387
2547 GGCAGTCTAGTGTCTTTTGGACTATATCTGAGGAGAGGTTATGACCCGCTGATG 2606
2388 AGATTTGGATGATCTGCGCTGGAATTTGCTTGGTAAATGCTCGGATTTGATCTGTATAG 2447
2607 TGCTAGGCTTAAGAGGATGATGAGGAGGAGGTTGTCGACAGACTGTCAAGATGCT 2666
2448 TGCCCGCTCGCTGGAAGAGGCAATCCAAATCTCAGCTGGAGAAATCCCTTTAGCGCAGAT 2507
2667 TGCTGCTGAGGAAATTCGATGAATACTCCACTGTTTACATAGATGAGGCGCTGCTGATCT 2726
2508 CTTAGCTGAGGCTTCGTGATGGGATCTGTGAGGTGGTGTGATGAGGAGGAGCT 2567
2727 TGT 2729
2568 TGT 2570

quence 7037, Application US/09815242
tent No. US20020061569A1

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zykend, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0

Q ID NO 7037
LENGTH: 2571
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(2571)
9-815-242-7037

ery Match 20.9%; Score 649.4; DB 9; Length 2571;
st Local Similarity 56.6%; Pred. No. 7.9e-173;
ches 1243; Conservative 0; Mismatches 946; Indels 6; Gaps 2;
426 TCTTATTAAGGTCAATCGTCTCAAGCTCTCAGAGTCAAGAGTCAAGAGTCAATTT 485
261 TATTAATCTACTTAATTTATGATAAATTCGACAGCAAAACCAAGATAAATTTATTC 320
486 GCGTGTGACCAAGTTGATTATGGTCTTCTTGAAGATTCTCAAAATCAGGGATTGTTGAA 545
321 GAGCGAATTTGTTTGTGTCAGCTTTAGAGAACGAGGAACGATCAGCGATATTTTGA 380
546 CGAAGTCGGTGTAGCGAGCGGAGGTAAGTCTGAGTGTGAGAGCTTCGTGGGAAAGA 605
381 AAGTTCGGTGTGAAAAGAAACAAATTTCCGACATTTAGAGGAGGAGCA 440
606 AGGGAAGAAAGTTGAGAGTGTCTCAGGGGACACAAATTTTCAAGCTTTAAAGACTTATGG 665
441 AAACGTGAACGATCAAAATGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 500
666 AAGAGATTTGGTTGAGCAAG 725
501 TTTAACCCTGCTGTCAGAAAGTGGCAAACTTGATCTGTAAATTTGGGCGGTGTAAGAAAT 560
726 TAGAAGAGTCGTGAGGATTTCTTCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 785
561 TCGTCAGAGCATTCAGATTATACAGCTGTTACCAAAATTAACCTGTGTAAATTTGGTGA 620
786 GCCAGAGTTGGTAAACAGAGCTGTGGTGAAGGTTTGTGAGAGGTTTGTGAGAGGTTTGTG 845
621 ACCAGGTGTAGGAAAACGGCGATTGTGAGAGGCTTGTGAGAGGCTTGTGAGAGGCTTGTG 900

QY 846 TGTGCCCAACAGTCTTACTGATGTGAGATTAAATTCGTGGACATGGGTGCGTTAGTTGC 905
DB 681 AGTGCAGAGAGGTTTGAATAAATTAACGTGCTTTCATATAGATATGGGGGCTTGAATTC 740
QY 906 TGGTGTCTAAATACCGAGAGAGTGTGAAGAAAGTGTGAATCTCTTTTGAAGAGAGTTGA 965
DB 741 TGGTGCAGAAATATCGTGTGAATTTGAAGAAAGTGTGAAGAAAGTGTGAATCTCAATGA 800
QY 966 GAGCGCTGAAGGCAAGTGTCTCTTTTATGATGAGATTCATTTGTTGTTCTTGTGCTGG 1025
DB 801 GAAAGAGAGTGCCTGTTATCTCTTTTATGAGAAATTCATATCTATGTTGCTGGCGGG 860
QY 1026 CAAACTGAAGGCTGATGATGAGTCAATCTCTTCAAGCCCATGTTAGCTAGAGGCA 1085
DB 861 TAAACCGATGTCGATGATGAGTCAATCTCTTCAAGCCCATGTTAGCTAGAGGCA 920
QY 1086 GCTTCGATGATGTCGATGATGAGTCAATCTCTTCAAGCCCATGTTAGCTAGAGGCA 1145
DB 921 ATTACATTCGCTGGTGCATCTCTTTAGATGAATATCGTCAATATATCGAAAAAGATGC 980
QY 1146 TGCCTTTGAGAGAGGTTTCCAAAGTCTATGTTGCGAGCCCAAGTGTGCTGACACAT 1205
DB 981 CGACTTGAACCGCTTTCCTCAAGGCTTTTGTGAGCAACCAAGTGTGAGAGATACCAT 1040
QY 1206 TAGTATCCTTAGAGGACTCAAGGAGAGATGATGAGGACATCATGCTGTGCGAATCCAAGA 1265
DB 1041 TCGGATCTTACGTGTTTGAAGAACGTTATGAATTCATCATCACTGATATCTGTA 1100
QY 1266 CAGAGCTCTTATAATGCTGCTCAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1325
DB 1101 CCCAGCAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1160
QY 1326 GGATAAAGCAATTCATTTGTTGATGAGGCTTGTGCGAATGCTGAGAGCTCCAGCTTGATG 1385
DB 1161 AGATAAAGCCATTCATTTGATGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1220
QY 1386 TCAACCTGAAGAGATTGATAACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1445
DB 1221 TAAACCTGAACCGCTTCTGATCGTCTTGAACGCTGCTGATTTATCCATTTAAATTTGA 1280
QY 1446 GCGCTTGAAG 1505
DB 1281 AGCGTTTCAAAAAG 1340
QY 1506 GCTTGTGATGCTGAG 1565
DB 1341 ATTGCTGAAAAG 1400
QY 1566 GAGAAATGATGAGATTCGAAGGCTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1625
DB 1401 AACGCTTCTGCGCTCTCAACATATTAACAGAGAGTTAGATCTGCAAAACCGAAGTGA 1460
QY 1626 GGAGGAGAGAGAGAGATGACCTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1685
DB 1461 ACAAGCTGCTGCGCGGGGTGATTTAGCGAAATGCTGAGTTGCAATATATGGCGCATCC 1520
QY 1686 AGAAGTGAATTCGAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTC 1745
DB 1521 TGATCTTGAAG 1580
QY 1746 CACAGAAACGTTGGGCGCTGAACACATGCTGAGGTTGTGAGCGCTTGGACAGAGAGATTC 1805
DB 1581 ACGCTATCGCGTCAAGATGAG 1640
QY 1806 AGTGAAG 1865
DB 1641 TGATCAAAATGATGAG 1700
QY 1866 TAAGCGGTTGTGGGACAGAGATCAAGCGGTAAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1925
DB 1701 TAAACGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1760
QY 1926 AAGGCGAGGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1985

1761 TCGTCGAGCTCTTTCCGATCCCTAAATCGCCCAATGGTCTCTTCTTGTTCTTAGGCCCAAC 1820
1986 TGGTGTGTGCAAACTGAGCTCGCAAGCTCTTCTGCTGAGCAGCTGTTTGATGATGAAAA 2045
1821 AGGTGTGGGAAACAGAGCTTTGCAAACTTTGGCTAAATCTTGTGTGATGATGAAAGA 1880
2046 COTCTTAGTTCGGATGTATATGTCGGAATATATGGAACAACACTCTGCTCTCGCCTCAT 2105
1881 TCGCATGTGGTGTATGATATGTCAGAGTTATGGAAGAACACACAGTGTCTCGTTAGT 1940
2106 TGGGCAACACAGGATGTTGTCACAGAGAGGTGGAACAATCACTAGGCTGTGAG 2165
1941 TGGTCGCTCCAGCTATGTCGGCTATGAGAGAGCGGTTATTTAACTGAAGCTGTTCG 2000
2166 GAGGCGACTTATGTCATCTCTTGTGAGTGGAGAGGCTCATGTTGCTGCTT 2225
2001 TCGTCGTCATATTCAGTGAUCTTATGATGAGTTGAAAGACACAGCATGTATT 2060
2226 CAACACTGTCTCCAAAGTTTGGATGATGGTGAATGACAGCGGCAAGGACGACAGT 2285
2061 CAATATCTTATACAAAGTGTGGATGATGGTCTTAACTGATGCTCAAGGTCTGACTGT 2120
2286 CGATTTCAGGAACCTCGGTGATATCATGACATCAACCTTGTGCTGACACCTCTTGC 2345
2121 GACTTCGCTAACACTGTGTTTATGATGACCTTAACTTGGGTTCTGATTTAATCCAAAG- 2179
2346 AGGGTAACTGGGAAAGTAAATGAAAGTGGCCGCGACTGTGTGATGCGGAGGTGAG 2405
2180 -GTAATAAGACGAAAGCTATAGCGAA- ---TGAAAGCTTAGTGTGATGATGAGT 2234
2406 GAAACACTTCAGACAGAGCTCTTGAAAGGCTGACAGAGTTGGTGTTCGACCCCT 2465
2235 CCAACATTTCCGCCAGAAATTCATCAACCGTATTGACGAAACCGGTGTTTCCATCACT 2294
2466 TTCACTGACAGCTTGCAGAAAGTGTGCTGCTCAATGAAAGAGTGTGCTGCGCT 2525
2295 TGGTAAAGAAATATCCGTGGATGCAAGTATCCAAATGAAAGCTTAGCAAAACGTAT 2354
2526 TGCTGAAAGAGAGTGTCTTTGGCAGTCACTGATGCTCTTTGGACTATATCTTTGGCAGA 2585
2355 GAAACTCGTGGATACGAATTTGGTGTGTACCGATGCTTTATAGACTTCATGCGGAAGT 2414
2586 GAGTTATCAACCGGTGATGTTGCTAGCTAGCCCTATAA 2620
2415 GGGATACGACCCCAATTTATGTTGTCAGCTCCATGCA 2449

ULT 13
10-329-670-1
sequence 1, Application US/10329670
Publication No. US20040018503A1
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
FILE REFERENCE: PB186P1
CURRENT APPLICATION NUMBER: US/10/329,670
CURRENT FILING DATE: 2002-12-24
PRIOR APPLICATION NUMBER: US 09/643,990
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: US 08/487,429
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: US 08/426,787
PRIOR FILING DATE: 1995-04-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1830121
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: misc_feature

LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)


```

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)

```

Query Match 20.9%; Score 649.4; DB 12; Length 1830121;

Best Local Similarity 56.6%; Pred. No. 7e-171; Mismatches 1243; Conservative 0; Matches 946; Indels 6; Gaps 2;

```

QY 426 TCTTATTAAGGTCATTTCGTCGCTCAAGCTCTCAGAGTCCAGAGTGATCACTCAITTT 485
Db 906639 TATTAATCTACTTAATTTATGCGATAAATTCACAGCAAAACCAAGATAAATTTATTC 906698
QY 486 GCGTGTGACCAAGTTGATTTATGGCTCTTCAAGATTCCTCAATCAGGGATTTGCGAA 545
Db 906699 GAGCGAAATTTGTTTTTTGTCAGACTTTAGAGAACGAGGAACGATCAGCGATATTTGAA 906758
QY 546 CGAAGTCGCTGTAGCGAGCGGAGGTAAGTCTCAGGTTCAGAGCTTCGTGGGAGAGA 605
Db 906759 AAAGTCGCGTGCAGAAAAGAAACAAATTCGCAAGCTATTACGCACTTAGGGGGGACA 906818
QY 606 AGGGAAGAAAGTTGAGAGTGTTCAGGGGACACAAATTTTCAAGCTTTAAAGACTTATGG 665
Db 906819 AAACGTGAACGATCAAAATGCGAAGAAAGACAGCAAGCGCTTGAAAAATATACGATTGA 906878
QY 666 AAGAGATTTGGTTGACGACGAGGAGCTTCTGATTCCTGATTTGCTGATGAGGAGAT 725
Db 906879 TTTAACCGCTCGTCAGAAAGTGGCAAACTTGTATCTGTAATTGGCGCTGATGAAGAAAT 906938
QY 726 TAGAAGAGTCGTGAGGATTTCTTCAGCAGAACGAGAACCAATCTCTGTCTTATTGGAGA 785
Db 906939 TCGTCAGCCATTCAAGTATTACACGCTGATCCAAAATTAACCTGTGTTAATTTGGTGA 906998
QY 786 GCCAGAGTTGGTAAAAACAGCTGTGGTTGAAGGTTTACACAAAGGATTTGAAAGGAGA 845
Db 906999 ACCAGGTGTAGGAAAAACGGGATTTGCGAAGGCTTGCCACAGCGCATCGTAAACGGGA 907058
QY 846 TGTGCCCAACAGTCTTACTGATGAGATTAATTTCTGTTGACATGGGTGCTGTAGTTGC 905
Db 907059 AGTGCCAGAAAGTTTGAATAAATAAAACGTGCTCTTCAATTAGATATGGGGCGCTTGAITGC 907118
QY 906 TGGTGTCTAAATACCGAGGAGAGTTTGAAGAAAGTTGAAATCTGTTTTTGAAGAGATTGA 965
Db 907119 TGGTCCGAATATCTGTTGATTTGAAGAACGTTTAAACAGTACTCAATGAACTTTC 907178
QY 966 GGACGCTGAAGCAAAAGTATCTCTTTTATTGATGAGATTCATTTTGGTTCTTGGTGTGG 1025

```

908259	TGCGATGGTGGCTATTGATATGTACAGATTTATGGAACACACAGTGTCTTCGCTTTAGT	9083118
2106	TGGGGCCACCCAGGGGTATGTGGTACACGAGGAAGTGCACAACTAACTGAGGGCTGTGAG	2165
908319	TGGTGGCGCTCCAGGCTATGTCGGCTATGAGAGGCGGTATTTAACTGAAGCTGTTTCG	908378
2166	GAGCGACCTTATTGTGTCTATCTCTTTGATGAAGTGGAGAGGCTCATGTTGCTGTCTT	2225
908379	TCGTCGTCCATATTCAGTGATCTTATTAGATGAAGTTGAAAAGCACACGCAGATGATTT	908438
2226	CAACACTCTGCTCCAGTTTGGATGATGATGGTGCATGACAGACGGGCAAGGACAGACAGT	2285
908439	CAATATCTTATTACAGTGTGGATGATGGTCGTTTTAACTGATGTCGAAGTCGTACTGT	908498
2286	CGATTTCAGAACTCGTGATATCATGACATCAACCTTTGGTGTCTGAACACCTCTTGC	2345
908499	GCATTTCCGTAACTGTGGTTATTATGACCTCTCACTTGGGTCTCGATTTAACTCCAAG-	908557
2346	AGGGCTAACTGGGAAAGTAAACATGGAAGTGGCCGGGACTGTGTGATCGGGAGGTGAG	2405
908558	-GTAATAAGACGAAAGCTATACGGAAA---TGAAGCCTTAGTGTGATGTCAGTGGTAAG	908612
2406	GAACACTTCAGACCGAGGCTCTGTGAACAGGCTTGACGAGATGTGGTGTTCGACCCGCT	2465
908613	CCAACTATTCGGCCGAGAAATTCATCAACGATATTACGAAACCGTGTATTCCATCCAAT	908672
2466	TTCACTAGCCAGTTGAGGAAAGTAGCTCGGCTTCAAAATGAAGAAGCGTGTCTGTCGGCT	2525
908673	TGGTAAAGAAATATCCGTGGGATTCGAAGTATCCAAATTAGAACGCTTAGCAAAACGTAT	908732
2526	TGCTGAAAGAGGAGTTGCTTTGGGCAGTCACTGATGCTGCTTTGGACTATATCTTTGGCAGA	2585
908733	GGAACTCGTGGATACGAATTGGTGTTTACCAATGCTTTATTAGACTTCATTGGCGAGT	908792
2586	GAGTTATGACCCGGGTGATGGTCTAGGCCTATAA	2620
908793	GGGATACGACCCAAATTTATGGTGCACGTCCAATTGA	908827

RESULT 14

```

US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleschmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)

```

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c

FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)

Query Match 20.9%; Score 649.4; DB 15; Length 1830121;
Best Local Similarity 56.6%; Pred. No. 7e-171;
atches 1243; Conservative 0; Mismatches 946; Indels 6; Gaps 2;

426 TCTATTAAAGTCATTCGTCGTCTCAAGCTGCTCAGAGTCACGAGGTGATCTCATTT 485
906639 TATTAACTTACTTAATTTATCGATAAATTCGCACAGCAAAACCAAGATAAATTTATTC 906698

486 GCCTTTGACCAAGTTGATTTATGGTCTTTTGAAGATTCCTCAAGATTCCTCAATCAGGGATTTGTTGAA 545
906699 GAGCAANTGTTTGTGTCAGCTTTAGAGAACGAGGAACGATCAGCENATTTTGA 906758

546 CGAGTCGTGTAGCGACGGCGAGGGTAAAGTCTGAGTTGAGAGGTTCTGTCGGAAGA 605
906759 AAAGTGCCTGCGAAAAAAGAACAAATTTCCGAAGCTATTTCAGCACATTTAGAGGGGACA 906818

606 AGGAAGAAAGTCGAGAGTGTCTCAGGGGACACAAATTTTCAAGCTTTAAAGCTATTGG 665
906819 AAAGTGAACATCAAAATGAGAGAACGAGAACAGCAAGCGCTTGAATAATATACGATTGA 906878

666 AAGAGATTTGTTGAGCAAGCAGGGAAGCTTGATCTCTGATTCGTGTCGTGATGAGAGAT 725
906879 TTTAACCGCTCGTCAGAAAGTGCACAACTTGATCTCTTAATTCGGCGTGTGAGAAAT 906938

726 TAGAAGAGTCGTGAGGATCTTTTCAGAGAGAACGAAGAACAACTCTGTCTTATTTGAGA 785
906939 TCGTCGAGCCATTCAAGTATTACAACTGTCGACCAAAATAACCTCTGTCTTAATTTGTTGA 906998

786 GCCAGAGTGTGTAACAGCTGTGTTGAAGTTTACACAAAGGATTTGAAAGGAGA 845
906999 ACCAGGTGTAGGAAGAACGGCGATTTGCAAGGCTTGGCACAGCGCATCTGAACGGCA 907058

846 TGTGCCCAACAGCTTTACTGATGTGAGATTAATTTGCTGACATGCGGTGCGTTAGTTGC 905
907059 AGTCCAGAGGCTTTGAAAAATAAACGTGTCTTTCAATAGATATGCGGGCTTTGATTGC 907118

906 TGGTGTCTAAATACCGAGAGAGTTTGAAGAGGTTGAAATCTGTTTTGAAAGAGTTGA 965
907119 TGGTGTGGAATATCTGTGTGTAATTTGAAGACGTTTAAAGCAGTACTCAATGAACCTTC 907178

966 GGACGCTGAAGGCAAGAGTATCTCTTTATTGATGAGATTCATTTGTTCTTGTGCTGG 1025
907179 GAAGAGAGAGTCTCGGTATTCCTTTTATTGACGAATTTCACTAATGTCGCGCGGG 907238

1026 CAAACTGAAGGGTCGATGAGTACGAGCTTAATCTGTTCAAGCCCACTGTTAGCTAGAGGCA 1085
907239 TAAACCCGATGTCGATGAGTGCAGGTAATTTGTTTAAACCAAGTTTGGCACGAGCGA 907298

1086 GCTTCGATGCAATTTGTCGTCACAACTGTTGAAGATACAGGAATATGTTGAGAAAGATGC 1145

907299 ATTACATTCGTCGGTGCACACTTACTTTAGATGAATATCTGTCATATATCGAAAAAGATGC 907358
1146 TGCCTTTGAGAGAGGTTCCAAACAAGTCTATGTTGCGAGCCCAAGTGTCCCTGACACCAT 1205
907359 CGCATTGAAACGCCGTTTCCAAAAAGTCTTTTGGACGAAACCAAGTGTAGAAAGATACCAT 907418
1206 TAGTATCTTTAGAGGACTCAAGAGAGATGATGAGGACATCATGTTGTCGCAATCCAGA 1265
907419 TGGCATCTTACGTTGTTTGAAGAGAGTTTGAATTCATCATCAGTGGATTTACTGA 907478
1266 CAGAGCTCTTATAAATGTCCTCAGCTGTCCTCTGTTACATAAATGTCGCTGCGCACTTACC 1325
907479 CCAGCAATTTGCTGCTGACCAACGCTTTACATGTTTATATTTCCGATCGCTAGTACC 907538
1326 GGNATAAGCAATTTGATTTGGTTGATGAGGCTTTGTCGGAATGTCAGAGTCCAGCTTGATAG 1385
907539 AGATAAAGCCATTTGATTTGATCGATGAAGCAGCGCTTAGCATTCGTATGGAATAGATTTC 907598
1386 TCAACCTGAAGAGATTGATAAATTTGAAGAGAGGATGACAGCTGGAAATTTGAACCTTCA 1445
907599 TAAACCTGAACCGCTTGATCGTCTTGAAGCTGCTATTTATCCAAATTAATAATTGGAACAACA 907658
1446 CGCTTTGGAAGAGAGATTAAGCCAGCAAGCTCGACTTATAGAGGTCGCGAAAGA 1505
907659 AGCTTTACAAAAAGAGAGAGCAAGCAAGTCGCAACGTTTAGAAATGTTAGAGAAAGA 907718
1506 GCTTGATGACCTGAGAGACAGCTTCAGCCTCTCAGGATGAATACAGAAAGGAGAAAGA 1565
907719 ATTGGCTGAAAAAGAACCGTGAATACGCCGAATCTGAAGAGTATGGAATCTGAAAAAGC 907778
1566 GAGAAATTGATGAGATTTGAAAGGCTTTAAACAGAAAAAGAGAGAGCTCATGTTTCTTTTCCGA 1625
907779 AACGCTTTCTGCTCTCAACATATTAAACAAGATTAGACTGCAAAAACCGAACTAGA 907838
1626 GGAGGAGAAACGAAGATATGACCTTTGCAAGAGCTGCTGATCTTAAGATATGCGCGCAATTCOA 1685
907839 ACAAGCTCTCGCGGGGTGATTTAGCGAAAAATGCTGAGTTGCAATATGCGCGCATCCC 907898
1686 AGAAGTGAATCTGCAATTTGCCCAATTTGAAGAACTTCTCTGAAGAGAAATGTCATGCT 1745
907899 TGATCTTGAAGAACACTTTGAGCAAGCTGAAACCGAGAGGAAAGAAATGACCGCTTTT 907958
1746 CACAGAAAAAGTTGGCGCTGAAACACATTTGTCAGAGTTGTCAGCCCTTGGACAGGATTTCC 1805
907959 ACCTATCGCTCAGATGAAGAAATCGCAGAGTGTCTTTTAAAGCCACAGGCACTTCC 908018
1806 AGTACGAGACTTTGGCCHAAATGAGAGAGAGGTTGATGCTGCTTGTCTGATGAGTTGCA 1865
908019 TGTATCAAAAATGATGGAAGCGCGGAAAGAAAACTCTTGGCTATGGAAGATGAACCTACA 908078
1866 TAAAGCGGTTGTCGACACAGAAATCAAGCGGTAATTCAGTCTTCTGAGGCAATTTCTAAGTCT 1925
908079 TAAACGAGTGTGTTGTCAGAGAGAGCGGTTGATCGGTCAGCAACCGCATTCGTCGTAG 908138
1926 AAGGCGAGACTTTGTTAGGCAACACAGCCAACTGGATCAATTTTATTTCTTGGACCAAC 1985
908139 TCGTCAGGCTCTTTCGATCCTTAATCGCCCAATTTGTTCTTTCTTCTTCTAGGCGCAAC 908198
1986 TGGTGTGCAAAAACCTGAGCTCGCCAGGCTCTTCTGAGCAGAGCTGTTTGTATGATGAAGAA 2045
908199 AGGTGTGGAAGAACAGAGCTTTGCAAAACTTTGGCTAAATTTCTTTTGTATGATGAAGA 908258
2046 CCTCTTAGTTCGATTTGATGTCGGAATATATGAAACAACTCTGCTCTCTCCGCTCAT 2105
908259 TCGCATGTCGCTATTTGATATGTCAGAGTTTATGGAAGAAACACAGTGTCTTCTGTTAGT 908318
2106 TGGGCGACCAACAGGATGATGTTGGTCACGAGAGAGGTCGCAACACTTAACGAGGCTGTGAG 2165
908319 TGGTGCCTCCTCAGGCTATGTCGGCTATGCAAGAGGCGGTTTATTTAACTGAAGCTGTTTCG 908378
2166 GAGGCGACCTTATTTGTCATCTCTTTGATGAGAGTGAAGAGGCTCATGTTGCTGCTCT 2225
908379 TCGTCGTCATTTTCAAGTGTCTTTATTAGTGAAGTTGAAAAAGCACACGAGATGTTAT 908438

2226 CAACACTCTCTCCAAAGTTTGGATGATGTTGATTCAGACGCGGCAAGCAGGACAGT 2285
 908439 CAATATCTTATTACAAAGTGTGGATGATGTTGATTTAACTGATGATGTTCAAGTCTGACTGT 908498
 2286 CGATTTCAGGACTCGGTGATATCATCAATCAACCTTGGTGTGAGACACCTCTCTTGC 2345
 908499 GGAATTCGCTTAACTGCTGTTATATGACCTTAACTTGGGTTCTGATTTAATCAAG- 908557
 2346 AGGCTAACTGGGAAAGTAACTTGAAGTGGCGGAGTGGCGGAGTGTGATGCGGAGGTGAG 2405
 908558 -GTATTAAGACGAAAGCTATAGGAAA-----TGAAAGCCTTATGATGTCAGTGTGTAAG 908612
 2406 GAAACACTTCAGACCGAGCTCTTGAACAGGCTTGACGAGATGTTGTTTCCACCCCTT 2465
 908613 CCAACATTTCCGCCCAAGATTCATCAACCGATTGACGAAACCGTGTGATTTCCATCCACT 908672
 2466 TTCACATGACCGTTGAGGAAAGTGTGCTGGCTTCAATGAAGACGTTGCTGTCCGGCT 2525
 908673 TGGTAAAGAAATATCCGTGCGATTGCAAGTATCAATTAAGACGCTTACGAAACGAT 908732
 2526 TGCTGAAGAGGAGTGTCTTTGGCAGTCACTGATGCTTTCGACTATATCTTTGGCAGA 2585
 908733 GGAACTCGTGGATACGAATGTTGTTACCGATGCTTTATTAGACTTCAATTGGCGAAGT 908792
 2586 GAGTTAGACCCGGTGTAGGTTGCTAGGCTATATA 2620
 908793 GGGATACGACCCCAATTTATGTCGACGCTCCATTGA 908827

LT 15
 0-369-493-33061
 Sequence 33061, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 PRIOR FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 Q. ID. NO 33061
 LENGTH: 2571
 TYPE: DNA
 ORGANISM: Xylella fastidiosa
 0-369-493-33061

ery Match 20.9%; Score 649; DB 12; Length 2571;
 st Local Similarity 55.4%; Pred. No. 1e-172;
 tches 1299; Conservative 0; Mismatches 1035; Indels 9; Gaps 2;
 396 TCCACCTGATGATATCCAGGAGTCTTAGTCTTATTAAGGTCATTCGTCGTCTCAGC 455
 228 TCAGACGGTTATGTCCTACCCAGCAATGAATGATGCTGTTGTTTCATCGGACTGATTA 287
 456 TGCTCAGAGTCAACGAGTGTACTCATTTGGCTGTGACCAAGTGTGATTTATGGGCTTCT 515
 288 GTTAGCGCAGCAGCATGGCGATCAGTTTATGGCCAGTGTGTTGTTGTTGCTGTGTGT 347
 516 TGAGGATTTCT--CAATCAGGATTTGTTGAACGAGTGTGTTAGCAGCGGCGAGGTT 572
 348 CGATGATAGTGGGGGCTGGCGGCTGGCTGTGCTGTGCTGTGCGGAAAGAAAGAGAT 407
 573 AAAGTCTGAGTTGAGAGCTTCGTGGGAAAGAGGAGGAGAAAGTGTGAGTGTCTCAGG 632
 408 TGAGGCTGCAATGATTAATTTGCTGGCGGTGAAACTGTCCAGACTGAGAAATGCTGAGGA 467

633 GGACACAAATTTTCAAGCTTTAAAGACTTATGGAAGAGATTTGGTTGAGCAAGCAGGAA 692
 468 GCAGCGTCAAGCGTTGGAAAAGTACACGATTTGATCTGACTGCGAGGGCTGAGAGTGGCAA 527
 693 GCTTGATCTCTGATTTGCTGCTGATGAGGAGATTTAGAAAGTCTGAGAGATTTCTTTCCAG 752
 528 GCTTGATCCGTTGATTTGCTGCTGATGAGAAATTCGCCGACGATACAGGTATTGCCAGG 587
 753 GAGAAGAGAAACATCTCTGCTTTATTGAGAGCCAGGATTTGGTAAAAACAGCTGTGGT 812
 588 CGGTACAAAAAATACTCTGCTGATTTGCTGAGCGGCTGCTGGTAAAAATGCCATTGT 647
 813 TGAAGCTTTAGCACAAAGATTTGAAAGGAGATTTGCCCAACAGTCTTACTGATGTGAG 872
 648 AGAAGCTTGGCTCAGCGCATTTGCAATGGTGAGGTTCCGAAAGGTTGCGCAGTAAGCG 707
 873 ATTAATTTCTGTTGACATGCTGCTTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 932
 708 TCTGCTCTCGCTGATTTGGGGGCTTGAATTTGCTGGGSCCAAGTTTCTGCTGAGTTTCA 767
 933 AGAAAGTTGAAATCTGTTTGAAGAAAGTTGAGGACGCTGAAAGGCAAGTGTCTCTT 992
 768 GGAGCGCTTGAAGGGGCTGCTTAAAGCATCTGCTAAGAAATGAGGGGCGGCTCATTTGTT 827
 993 TATTGATGAGATTTCAATTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1052
 828 CATTGACGAGCTGCATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 887
 1053 TAATCTGTTCAAGCCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1112
 888 CAATATGCTCAAGCGGCTTAGCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 947
 1113 TGAAGAAATACAGGAAATATGTTGAGAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1172
 948 GGATGAGTATCGCAAGTACATTGAGAAAGGATGCGCGCTTGGAAACGCTTTTCCAGAAAGT 1007
 1173 CTATGTTGCGGAGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1232
 1008 GTTGTGTTGGGAGCGGAGTGGAGGATCCATCGCAATTTCTGCTGCTGCTGCTGCTGCTGCTG 1067
 1233 GTATGAGGGAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1292
 1068 GTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1127
 1293 GTCTG 1352
 1128 GTCTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1187
 1353 GGCTTGTGCGAAATGCTGAGGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1412
 1188 GGCTGCCAGTCTGATTTGATGGAATTTGCTCCAGCGGAGAGGCTTGTGCTGTTGGA 1247
 1413 AAGGAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1472
 1248 CGCTG 1307
 1473 CAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1532
 1308 GAGCAGCAGGCTTACCGGATCTTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1367
 1533 GCCTCTCAGATGAAATACAGAAAGGAGAAAGAGAAATTTGATGAGATTTGAAAGGCTTAA 1592
 1368 CGATCTGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1427
 1593 ACAGAAAGAGAGAGCTCATGTTTCTTTCAGGAGGAGGAGAAAGAGAGATATGACCTTGC 1652
 1428 GGACTCGATCGAGCAGCAAGCTTGTGTTTGGAGCGCGCAGCGCTGCTGCTGCTGCTGCTGCTG 1487
 1653 AAGAGCT 1712
 1488 CAAGATGAGCGAGATTCAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1547

1713	TGAAGGAAC	TTC	TCTTCTG	AAGAGAA	TGTGATCT	CA	CAGAA	ACGTTGG	CCCTG	AACACAT	1777
1548	AAGTCAG	CGCGAA	CAGCAC	CGATTTT	ACTTTTGGT	TCAGG	AAGAA	AGTGACT	CTGCTGAG	GAGAT	1607
1773	TGCTGAG	CTTG	TGAGCCGTT	TGGACAG	GGATTTCCAGT	GCAGAC	ACTTGC	CCAAA	TGAGAA	1832	
1608	TGCCGAG	TAGT	TAGTCGTT	TGGACTG	GTATTC	CA	GTGAC	AGATGCTT	GAGGCG	1667	
1833	GGAGAG	TTGAT	TGGTCTT	GTGATAG	TTGCATA	AAG	CGGTTG	TGG	ACAGATCA	1892	
1668	CGATAAG	TTGTT	TGGGTAT	TGGAAGCT	GTATCT	AGT	CTCAG	CTGTG	TTGCTCA	1727	
1893	GGTAANT	CCAGTTT	CTCAG	GCCTTCT	TAAGGTCA	AGG	CGGAC	CTTGTG	TAGGCG	1952	
1728	GATCAAG	GTGAT	TCCGAT	GCGGTG	CCCGCTT	CG	CTACT	TGGTTT	GTCTGAT	1787	
1953	GCCAACT	CGATCA	TTCCTT	ATCTTGC	ACCACT	GTGT	TGG	CAAACT	GTAGCT	2012	
1788	ACCGAG	CGGCTCG	TTCCTT	CTGGT	CCGAC	CGGTG	TGTA	AGACT	GAGTTAT	1847	
2013	GGCTCTT	CTCAG	CAGCTG	TTTGTAT	GTATG	ATG	AAACCT	CTTAGT	TCGAT	2072	
1848	GGCGT	CTGCTCAA	TTTTT	GTTC	CA	AGTCA	AGATG	CAAT	TG	1907	
2073	ATATAT	GAA	CAACACT	CTCTCT	CTCG	CTCAT	TGG	GGCAC	CA	2132	
1908	GTTCA	TG	AAAAA	CA	TTCTGT	GG	CGCTG	ATTTG	TGG	1967	
2133	CGAG	AGGTG	CGACA	CTAACT	GAGGCT	GTG	AGG	CGG	CA	2192	
1968	TGAG	AAAGG	AGTTAT	CTTACT	GAAT	TGGT	CGCA	GTG	CG	2027	
2193	TGAT	CAAGT	GAG	AA	GCTCA	TGTC	CTCTCA	CACT	CTG	2252	
2028	GGAT	GAAGT	GGA	AGCG	CA	TAGT	GATG	TGTT	CA	2087	
2253	TGGT	CGAT	TG	ACA	GGGCA	AGG	CAG	CACT	CG	2312	
2088	TG	ACGCTTG	ACTG	ATGGT	CAG	GGCGT	ACTG	TG	ATTT	2147	
2313	GACAT	CAAA	CTTG	TGCT	GAA	CAC	CTCT	TG	AGG	2372	
2148	GACAT	CAAACT	TAG	TTTCA	CAT	CAG	ATCCA	AGAACT	CT	2207	
2373	AGTGG	CCCC	----	GG	CACT	GTG	TATCG	GGAGGT	GAG	2426	
2208	GTAT	ACG	CAAT	G	AGCG	CACT	GAT	TGG	GGT	2267	
2427	CTT	GAA	CAGGCT	TGA	CGAGAT	TGGT	GT	TG	CA	2486	
2268	CA	TAA	CAGGTT	GGAT	ATTTG	TGTTT	CA	CCGTT	G	2327	
2487	AGT	AGCT	CGGCTT	CA	ATG	AA	ACGCTG	CT	CG	2546	
2328	GATT	CG	CGGAT	CTCA	CTG	CGG	GGCT	CT	G	2387	
2547	GG	CAGT	CACT	GTGCT	GTG	ACT	TATCT	TG	CA	2606	
2388	GA	TTTGG	ATG	ATCGT	GGCT	TG	AA	TG	TC	2447	
2607	TGCT	TAG	CCCTTAT	AG	GAGAT	TGGAT	TG	AG	AG	2666	
2448	TGCC	CGCCG	CGCT	GA	GAGAG	CA	ATCC	AA	TCT	2507	
2667	TGT	CGT	GAG	GA	ATCG	AT	CA	AACT	CT	2726	
2508	CTT	AGCT	TGG	ACG	TTT	CTG	TAT	CA	TAG	2767	
2727	TGT	2729	---	---	---	---	---	---	---	2567	
2568	TGT	2570	---	---	---	---	---	---	---	---	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

rotein - nucleic search, using frame_plus_p2n model

on: February 13, 2004, 01:41:41 ; Search time 142 Seconds
(without alignments)
2831.690 Million cell updates/sec

e: US-09-812-350-17

ect score: 4550
ence: 1 MPEKTHKTNTATHEL.....VKMREIEDDDNEEMIED 911

ing table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

ched: 569978 seqs, 220691566 residues

1 number of hits satisfying chosen parameters: 1139956

um DB seq length: 0

um DB seq length: 2000000000

-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

and line parameters:

EL=frame+p2n.model -DEV=xlpl

cg2 1/USPTO spool_p/US09812350/runat 10022004 101516 13378/app_query.fasta_1.1095

Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCHI=0

PEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi

T=45 -DLOCAL=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15

B=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -NCPUS=6 -ICPU=3

TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

PEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

base : Issued Patents NA: *
1: /cg2 6/ptodata/2/ina/5A_COMB.seq: *
2: /cg2 6/ptodata/2/ina/5B_COMB.seq: *
3: /cg2 6/ptodata/2/ina/6A_COMB.seq: *
4: /cg2 6/ptodata/2/ina/6B_COMB.seq: *
5: /cg2 6/ptodata/2/ina/PTCUS_COMB.seq: *
6: /cg2 6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

ID	DB	Length	Query % Match	Description	
				Score	
2194	4	2574	48.2	US-08-887-534A-46	Sequence 46, Appl
2194	4	2574	48.2	US-09-527-431-46	Sequence 46, Appl
2173.5	4	2628	47.8	US-09-107-532A-883	Sequence 883, App
2162.5	4	1830121	47.5	US-09-557-884-1	Sequence 1, Appl
2162.5	4	1830121	47.5	US-09-643-990A-1	Sequence 1, Appl
2158.5	3	4403765	47.4	US-03-103-840A-2	Sequence 2, Appl
2158.5	3	4411529	47.4	US-09-103-840A-1	Sequence 1, Appl
2140	4	2592	47.0	US-09-328-352-2604	Sequence 2604, Ap
2133	4	2580	46.9	US-09-199-637A-280	Sequence 280, App
2133	4	2970	46.9	US-09-199-637A-272	Sequence 272, App
2133	4	42235	46.9	US-09-199-637A-1	Sequence 1, Appl
2132	4	2580	46.9	US-09-252-991A-4842	Sequence 4842, Ap

Sequence 4897, Ap
Sequence 1142, Ap
Sequence 1, Appli
Sequence 1244, Ap
Sequence 1, Appli
Sequence 5922, Ap
Sequence 3, Appli
Sequence 3, Appli
Sequence 2055, Ap
Sequence 1, Appli
Sequence 123, App
Sequence 5965, Ap
Sequence 18, Appl
Sequence 2033, Ap
Sequence 168, App
Sequence 6033, Ap
Sequence 5908, Ap
Sequence 684, App
Sequence 727, App
Sequence 661, App
Sequence 14, Appl
Sequence 1, Appli
Sequence 14627, A
Sequence 15131, A
Sequence 3366, Ap
Sequence 84, Appl
Sequence 8976, Ap
Sequence 2353, Ap
Sequence 11, Appl
Sequence 276, App

US-09-252-991A-4897
US-09-134-001C-1142
US-09-198-452A-1
US-08-249-380-1
US-09-134-001C-1244
US-09-040-843-1
US-09-621-855-1
US-09-252-991A-5922
US-09-040-843-3
US-09-621-855-3
US-09-107-532A-2055
US-08-545-528D-1
US-08-961-527-123
US-09-252-991A-5965
US-09-724-623-18
US-09-107-532A-2033
US-08-961-527-168
US-09-252-991A-6033
US-09-252-991A-5908
US-09-252-991A-684
US-09-252-991A-727
US-09-221-017B-661
US-09-724-623-14
US-08-923-511-1
US-09-416-874A-1
US-09-252-991A-14627
US-09-252-991A-15131
US-09-328-352-3366
US-08-961-527-84
US-09-252-991A-8976
US-09-328-352-2353
US-09-453-702B-11
US-09-199-637A-276

ALIGNMENTS

RESULT 1
US-08-887-534A-46
; Sequence 46, Application US/08887534A
; Patent No. 6455323

GENERAL INFORMATION:

APPLICANT: Holden, David W.

TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive/6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/887,534A

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Heien

REGISTRATION NUMBER: 33,547

REFERENCE/DOCKET NUMBER: 28341/33996

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

TELEX: (312) 474-6600

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:

LENGTH: 2574 base pairs

TYPE: nucleic acid

STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "DNA (genomic) (p10b30)"
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..2571
 88-887-534A-46

Alignment Scores:
 J. No.: 9.01e-214 Length: 2574
 Score: 2194.00 Matches: 447
 Percent Similarity: 69.56% Conservative: 154
 Mismatches: 239
 Indels: 24
 Gaps: 9

09-812-350-17 (1-911) x US-08-887-534A-46 (1-2574)

1 MetAsnProGluLysPheThrHisGlyThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 1 ATGGCTCGGATCTCTTAATAAATTCAGCTTCTTGGCGATGCCCAATCACTT 60
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuSer 40
 61 GCATCGGGCAGCAGCAACCAATTATCGAACCTTTCATTTAATGAGCGCCCTGCTGAAT 120
 41 AspProThrGlyPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
 121 CAGCAAGGGGTCGGTTAGTCCCTTTAATCAATCCGCT--GGCATAAATGCTGCCAG 177
 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
 178 ---TTGGCCACAGATATCAATCAGGCATTAATATGTTTACCGCAGGTTGAAGGTACTGGT 234
 81 AspAspIleProAlaSerSerLeuIleLysValIleArgAlaGlnAlaAlaGln 100
 235 GGTGATGTCAGCCATCACAGGATCTGGTGGCGTCTTAACTTTCTTCCGACACAGTGGCG 294
 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuMetGlyLeuLeuGluAsp 120
 295 CAAAACGCTGGTGATACTTATCTCTCAGAACTGCTTCTTCTGGCGGCATCTGAGTCT 354
 121 ---SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
 355 CGCGCCACCGTGGCGGCATCTCTGAAAGCAGCAGCGCGCACCCGCAACATTAATCTCAA 414
 140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159
 415 GCGATTGAAACAATGCGTGA-----GGTGAAGCGTGACGATCAAGTGCTGAGAC 468
 160 AsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLys 177
 469 CAACGTCAGGCTTTGAAAAAATATACCATCGACTTACCAGAACGAGCGCAACAGCGCAAA 528
 178 LeuAspProValIleGlyArgAspGluGluIleArgValValAlaGlyLeuSerArg 197
 529 CTCGATCCGCTGATGCTCGATGAAGAATTCGCGTACCATTACAGGTGCGCAACGT 598
 198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217
 589 CGTACTAAATAACCCGGTACTGATTGGTGAACCCCGCTCGGTAATACTGCCATCGTT 648
 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 649 GAAGGTCTGGCGCAGCGTATTATCAACGGCGAAGTCGCGAAGGGTTGAAAGGCCCGG 708
 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGlu 257
 709 GTACTGGCGCTGGATATGGGCGCGCTGGTGGCGGAAATATCGCGTGAAGTTGAA 768
 258 GluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPhe 277

769 GAACGCTTTAAAGCGGTGCTTAACGATCTTGCCAAACAGGAGCAAGCTCATCTCTATTT 828
 278 IleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAla 297
 829 ATCGAGCAATTACATACCATGTCGCGGGGTAAGCGGATGGCGCAATGAGCCCGGA 888
 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeu 317
 889 AACATGCTGAAACCGCGCTGGCGGTGCTGAATTGCACTGCTAGGTGCCACGAGCTT 948
 318 GluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnVal 337
 949 GACGAATATCGCCAGTACATTTGAAAAAGATGTCGCGTGGAAACGTCGTTTCCAGAAAGTG 1008
 338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLys 357
 1009 TTTGTTGCCGAGCTTCTGTTGAAGATACCATTTGCGATCTGCGTGGCTGAAAGACGT 1068
 358 TyrGluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377
 1069 TAGCAATTGCAACCATGTCGCAAAATTACTACCGCGCAATTGTTGACGCGCGACGTTG 1128
 378 SerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 1129 TCTCATCGCTACATTTGTCGACCGTCACTGCGGATAAGCCATCGACTGATCATGATAA 1188
 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGlu 417
 1189 GCAGCATCCAGCATTCGTATGTCAGATTGACTCAAAACAGAAAGACTCGACCGACTCGAT 1248
 418 ArgLysArgMetGlnLeuGluIleGluIleHisAlaLeuGluArgGluLysAspLysAla 437
 1249 CGTCGTATCATCCAGCTCAAACTGGAAACACAGCGGTAAATGAAAGAGTCTGATCAAGCC 1308
 438 SerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGln 457
 1309 AGTAAACAGCTCTCGATATGCTCAACGAAAGCTGACCGCAAGAACGTCGACTACTCC 1368
 458 ProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLys 477
 1369 GAGTTAGAAGAGAGTGGAAAGCAGAGAGGATCGCTTCTGTTACGACAGACCATTA 1428
 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAla 497
 1429 GCGGAAGTGGAAACAGCGCAAAATCGCTATTGAACAGGCTCGCGCTGTGGGGACCTGCGC 1488
 498 ArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGlu-----SerAlaIle 514
 1489 CGGATGCTGAACTGCAATACGCAAAATCCGGAACTGGAAAGCAACTCGAAGCCGCA 1548
 515 AlaGlnLeuGluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyPro 534
 1549 ACGCAGCTCGAAGGCAAACTATGCGT-----CTGTTGCGTAATAAAGTGAACCGAC 1599
 535 GluHisIleAlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGln 554
 1600 GCCGAATATGCTGAAGTCTCGCGTGGAGCGGGATTCGCGTTTCTCGCATGATGAA 1659
 555 AsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGln 574
 1660 AGCAGCGCGCAAAACTCTCGGTATGGAGCAGCAAGACTGCACCATCGGTAAATGGTCAG 1719
 575 AsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArg 594
 1720 AACGAAGCGGTGATGCGGTATCTAAACGCTATTCTGCTAGCCGCGGGCTGGCGAT 1779
 595 AlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGlu 614
 1780 CCAATATCCCGCATTTGTTTCAATCTCTGTTCTCGGCCCACTGTTGTGGGGGAAACAGAG 1839
 615 LeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAsp 634
 1840 CTTTGAAGCGCTGGCGAACTTTATGTTGATAGCAGCGAGCGATGGTCCGTATCGAT 1899

635 MetSerGluTyrMetGluGlnHisSerValSerArgLeuLeuGlyAlaProGlyTyr 554
 1900 ATCTCCGAGTTTATGAGAAACACTCGGTCTCTGTTGGTGGCTCGGGATAT 1959
 655 ValGlyHisGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysVal 674
 1960 GTCCGTTATGAGAGTGCTACCTACCTGACCGAGCGGTGGTCTGCTCGTATTCGGTC 2019
 675 IleuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuGlnVal 694
 2020 ATCTCTGTCGATGAAGTGAAGAAAGGCGATCCGATGCTTCAACATCTGTTCAGGTA 2079
 695 LeuAspGluValGluLeuThrAspGlyGlnGlyValArgThrValAspPheArgSerVal 714
 2080 CTGGATGATGGGCTCTGACTGACGGCAGGAGAGACGTCGATTCGTTATACGGTC 2139
 715 IleileMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysVal 734
 2140 GTCATTTATGACCTTAACCTCGTTCGGAT---CTGATTCAGGAACGCTTCGGTGAACGTG 2196
 735 ThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGlu 754
 2197 GATTATGCGCACATGAAGAGCGTGGTCTGCTGCTGTAAGCCATAACTTCGCTCGGAA 2256
 755 LeuLeuAsnArgLeuAspGluLeuValValPheAspProLeuSerHisAspGlnLeuArg 774
 2257 TTTCAATTAACGTTATGATGAGTGGTGGTCTTCATCCGCTGGTGAACAGCACATGGC 2316
 775 LysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAla 794
 2317 TCGATTGCGCAGATTCAGTTGAAACGCTCTACAAACGCTGGAAGACGTTGTTATGAA 2376
 795 LeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyr 814
 2377 ATCCATTTCTGACGAGGCGCTGAAACCTGCTGAGCGAGAACGGTTACGATCGGCTAT 2436
 815 GlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMet 834
 2437 GGTGACGCTCTGAAACGTCGAATTCAGCAGCAGATGAAACCCGCTGGCAGCAA 2496
 835 -----ValValArgGluLeuAspGluAsnSer 844
 2497 ATACTGTCTGGTGAATGGTTCGGGTAAAGTATTGCTGCTGGAAGTTAATGAAGACCGG 2556
 845 ThrValTyrIle 848
 2557 ATTGTCGCGTC 2568

FILE 2
 9-527-431-46
 quence 46, Application US/09527431
 tent No. 6485899
 GENERAL INFORMATION:
 APPLICANT: Holden, David W.
 TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 STREET: 233 South Wacker Drive/6300 Sears Tower
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/527,431
 FILING DATE:
 CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/887,534
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Rin-Laures, Li-Hsien
 REGISTRATION NUMBER: 33,547
 REFERENCE/DOCKET NUMBER: 28341/33996
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 474-6300
 TELEFAX: (312) 474-0448
 TELEX: (312) 474-6600
 INFORMATION FOR SEQ ID NO: 46:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2574 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "DNA (genomic) (p10b30)"
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..2571
 US-09-527-431-46

Alignment Scores:
 Pred. No.: 9, 01e-214 Length: 2574
 Score: 2194.00 Matches: 447
 Percent Similarity: 69.56% Conservative: 154
 Best Local Similarity: 51.74% Mismatches: 239
 Query Match: 48.22% Indels: 24
 DB: 4 Gaps: 9

US-09-812-350-17 (1-911) x US-09-527-431-46 (1-2574)

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 Db 1 ATGCGTCTGGATCGTCTTACTATAAATTCAGCTGCTCTTCCGATGCCAATCACTT 60
 QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLysSer 40
 Db 61 GCATCTGGGCGCAGCAACCAATTTATCGAACCACTTCATTTAATGAGCGCCCTGCTGAAT 120
 QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
 Db 121 CAGGAAGGGGTTCCGTTAGTCTTATTAACATCCGCT---GGCATAATGTGGCCAG 177
 QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
 Db 178 ---TTGCGCACAGATATCAATCAGGCATTAATCGTTTACCGCAGGTTGAAGGTACTGCT 234
 QY 81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGln 100
 Db 235 GGTGATGTCAGCCATCAGGATCTGGTGGCGGCTTCTTAATCTTCGCGACAGGTGGCG 294
 QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAsp 120
 Db 295 CAAAAACGTGTGTATTAATCTCTCAGAACTGTTCTGTCGGCGCACITGACTCT 354
 QY 121 ---SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
 Db 355 CGCGGCACCGTGGCGGCATCTCTGAAAGCAGCAGGGCGACCCGCCCAACATTAATCA 414
 QY 140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159
 Db 415 GCGATTGAACAAATGCGTGA---GGTGAAGCGTGAACCGATCAAGTGTCTGAAGAC 468
 QY 160 AsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLys 177
 Db 469 CAACGTGAGGCTTTGAAAAAATATACCATCGACCTTACCGACGAGCCGACAGGCGCAA 528
 QY 178 LeuAspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArg 197
 Db 529 CTCGATCCGGTGTGTTGGTCTGTGTAAGAAATTCGCGGTACCATTCAGTGTCTGCAACGT 588

555	AsnGluLysGluArgLeuLeuLeuGlyLeuAlaAspArgLeuHisLysArgValValGlyGln	574
QY	AsnGluLysGluArgLeuLeuLeuGlyLeuAlaAspArgLeuHisLysArgValValGlyGln	
1660	AGCGAGCGCGCAAAACTGCTGCCTATGAGGACAGAACTGCACCATCGCGTAATTTGGTCAG	1719
Db	AGCGAGCGCGCAAAACTGCTGCCTATGAGGACAGAACTGCACCATCGCGTAATTTGGTCAG	
575	AsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArg	594
QY	AsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArg	
1720	AACGAAGCGGTGATGCGGTATCTAACCGTATTTCTGCTAGCCGTCGGGGCTGGCGGAT	1779
Db	AACGAAGCGGTGATGCGGTATCTAACCGTATTTCTGCTAGCCGTCGGGGCTGGCGGAT	
595	AlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGlu	614
QY	AlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGlu	
1780	CCAAATCGCCCGATTGGTTTCATTCCTGTTCTCTCGGCCCAACTGSGTGTGGGAAAACAGAG	1839
Db	CCAAATCGCCCGATTGGTTTCATTCCTGTTCTCTCGGCCCAACTGSGTGTGGGAAAACAGAG	
615	LeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAsp	634
QY	LeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAsp	
1840	CTTTGTAAAGCGCTGCGCAACTTTATGTTTGTATGATGACGAGCGATGGTCCGTATCCGAT	1899
Db	CTTTGTAAAGCGCTGCGCAACTTTATGTTTGTATGATGACGAGCGATGGTCCGTATCCGAT	
635	MetSerGluTyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyr	654
QY	MetSerGluTyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyr	
1900	ATGTCGAGGTTTATGAGAAACACTCGGTGTCGTGTTGGTGGTGGCCCTCCGGGATAT	1959
Db	ATGTCGAGGTTTATGAGAAACACTCGGTGTCGTGTTGGTGGTGGCCCTCCGGGATAT	
655	ValGlyHisGluGluGlyGlnLeuThrGluAlaValArgArgArgProTyrCysVal	674
QY	ValGlyHisGluGluGlyGlnLeuThrGluAlaValArgArgArgProTyrCysVal	
1960	GTCCGGTTATGAAGAAGTGCGTACCTACCGAAGCGTGGCTCGTCCGTATCCGCT	2019
Db	GTCCGGTTATGAAGAAGTGCGTACCTACCGAAGCGTGGCTCGTCCGTATCCGCT	
675	IleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnVal	694
QY	IleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnVal	
2020	ATCCTCTCGATGAAGTGAAAAGCGCATCCCGATGTCTTCAACATTCCTGTTCAGGTA	2079
Db	ATCCTCTCGATGAAGTGAAAAGCGCATCCCGATGTCTTCAACATTCCTGTTCAGGTA	
695	LeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerVal	714
QY	LeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerVal	
2080	CTGGATGATGGGCGTCTGACTGACGGCGAAGGAGAACGGTCCACTTCGGTAATACGGCT	2139
Db	CTGGATGATGGGCGTCTGACTGACGGCGAAGGAGAACGGTCCACTTCGGTAATACGGCT	
715	IleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysVal	734
QY	IleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysVal	
2140	GTCAATTATGACCTTAACCTCGGTTCCGAT---CTGATTCAGGAACGCTTCGGTGAACCTG	2196
Db	GTCAATTATGACCTTAACCTCGGTTCCGAT---CTGATTCAGGAACGCTTCGGTGAACCTG	
735	ThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGlu	754
QY	ThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGlu	
2197	GATTATCGGCACATGAAGAGCTGGTGTCTCGGTGTGTGTGAAGCCATACTTCGCTCCGGAA	2256
Db	GATTATCGGCACATGAAGAGCTGGTGTCTCGGTGTGTGTGAAGCCATACTTCGCTCCGGAA	
755	LeuLeuAsnArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArg	774
QY	LeuLeuAsnArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArg	
2257	TTCAATTAACTGATCATGAAGTGGTGGTCTTCATCCGCTGGGTGAACACATTCGCC	2316
Db	TTCAATTAACTGATCATGAAGTGGTGGTCTTCATCCGCTGGGTGAACACATTCGCC	
775	LysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAla	794
QY	LysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAla	
2317	TCGATTGGCGAGATTTCAGTTGAAACGCTGTACAACGCTCTCGAAGAACGTTGTTATGAA	2376
Db	TCGATTGGCGAGATTTCAGTTGAAACGCTGTACAACGCTCTCGAAGAACGTTGTTATGAA	
795	LeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyr	814
QY	LeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyr	
2377	ATCCACATTTCTGACAGCGCTGAAACCTGCTGAGCGAGAACCGTTTACGATCCGGTCTAT	2436
Db	ATCCACATTTCTGACAGCGCTGAAACCTGCTGAGCGAGAACCGTTTACGATCCGGTCTAT	
815	GlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMet	834
QY	GlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMet	
2437	GGTGCACGCTCTGAACGTCGAATTCAGCAGCAGATCGAAAACCCGCTGGCCACAGCAA	2496
Db	GGTGCACGCTCTGAACGTCGAATTCAGCAGCAGATCGAAAACCCGCTGGCCACAGCAA	
835	-----ValValArgGluGluIleAspGluAsnSer	844
QY	-----ValValArgGluGluIleAspGluAsnSer	
2497	ATACTGCTCTGGTGAATTGGTTCGGGTAAAGTGATTTCGCTGGAGCTTAATGAAGACCG	2556
Db	ATACTGCTCTGGTGAATTGGTTCGGGTAAAGTGATTTCGCTGGAGCTTAATGAAGACCG	
845	ThrValTyrIle	848
QY	ThrValTyrIle	
2557	ATTGTCGCGGTC	2568
Db	ATTGTCGCGGTC	

1. TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

1282 ATGCAGTTGGAAATCGAAGAAGCTGCACTGAAAAAGAAATCAGATGATGCAAGTAAAAA 1341

```

Db      2422  GATGAAGCCAAAACGTTGGATTGCTGAAAAATGCTTATGAACCCAGCATATGTCGCAAGACCA 2481
QY      819   IleArgTIpMetGlulYslyslYsValValThrGlulLeuSerlyslMetValValA:ggLu 838
Db      2482  TTCAAAACGATTATCTACTAAGGAAGTAGAAACACCACTAGCTAAAGAAATCGTTGCAGGA 2541
QY      839   GluIleAspGluAnsSerThrValTyrlleAspAlaGlyAlaGlyAspIleuValTyArg 858
Db      2542  CATGTTTATGCTCTAAATCAAAAGTAACGATTACTTTACTTTGATGAGCAACTTCTATTTCAAA 2601
QY      859   Valglu 860
Db      2602  ACAGAA 2607

RESULT 4
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PBI86P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
; US-09-557-884-1

Alignment Scores:
Pred. No.: 5,73e-206 Length: 1830121
Score: 2162.50 Matches: 444
Percent Similarity: 69.9% Conservative: 153
Best Local Similarity: 52.05% Mismatches: 245
Query Match: 47.53% Indels: 11
DB: Gaps: 8

US-09-812-350-17 (1-911) x US-09-557-884-1 (1-1830121)

QY      1   MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
Db      906379  ATGAACATTGAAAAATTTACGACAAAATTCGAAGAGCGTTAAGCGAACGACCAATCACTT 90643

```


21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
||| : : : : :
906439 GCTATCGGAGGAGCAATCAATTTATTGTAACCTGTGCAATTTATTGACCGCACTTTAAAT 906498
||| : : : : :
41 AspProThrGly---IlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAla 59
||| : : : : :
906499 CAACAGGGCGCTCTATTGCGCA-----ATTTCAGCAGCAAGTGGCGTAATGTGGCT 906552
||| : : : : :
60 GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 79
||| : : : : :
906553 ---TTATTGCGTAATCAATTAAGACAGACNACTAATAATATGCCCAAGTATGGCAAT 906609
||| : : : : :
80 ProAspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAla 99
||| : : : : :
906610 GCGCGTGATGTACAACTTTACGCCAGCTTATTACTACTTAAATTTATGCGATAAATTC 906669
||| : : : : :
100 GlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGlu 119
||| : : : : :
906670 GCACAGCAAAACCAAGATAAATTTATTGCGCGAATTTGTTTTCGACGCTTTAGAA 906729
||| : : : : :
120 Asp---SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
||| : : : : :
906730 GAACGAGGACATCAGCATATTTTGAAGAGTGGCGTGGCAAAAGAACAAATTCG 906789
||| : : : : :
139 SerGluValGluLysLeuArgGlyLysGluGlyLysLysValGluSerAlaSerGlyAsp 158
||| : : : : :
906790 CAAGCTATTTCAGCATATTAGGG-----GGACAAACGTTGAACGATCAAAATGCAGAA 906843
||| : : : : :
159 ThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----Gly 176
||| : : : : :
906844 GAAACGACAGACGCGCTTGAATAAATATACGATTGATTTAAACGCTCGTGCAAGAGTGGC 906903
||| : : : : :
177 LysLeuAspProValIleGlyArgAspGluGluIleArgValValArgIleLeuSer 196
||| : : : : :
906904 AAACCTTGATCTGTATTTGGGCGTATGAGAAATTCGTGAGCAATTCAGATATTACAA 906963
||| : : : : :
197 ArgArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaVal 216
||| : : : : :
906964 CGTCGTACCAAAATAACCTGTGTAAATTTGGTGAACAGGTGTAGGGAACACGCGGATT 907023
||| : : : : :
217 ValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspVal 236
||| : : : : :
907024 GTGGAAGCGCTTGGCACGCGCATCTAAACGCGAGTGCAGAGGTTTGAATAAATAA 907083
||| : : : : :
237 ArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyClyPhe 256
||| : : : : :
907084 CGTGTGCTTTCATTAGATATGGGGCGTGTATTGCTGGTGGCAATATCGTGTGAATTT 907143
||| : : : : :
257 GluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeu 276
||| : : : : :
907144 GAAAGAACGTTTAAAGACGAGTACTCAATGAACCTTCGAAAGAGAGGTCGCGTTATCCTC 907203
||| : : : : :
277 PheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAla 296
||| : : : : :
907204 TTTATGACGAATTCATCTAGTGTGGCGGGTAAACCGGATGGTGGCGATGCG 907263
||| : : : : :
297 AlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThr 316
||| : : : : :
907264 GGTAAATTTGTTAAACCAAGTTTGGCACAGCGCAATTTACATTGGTGGTGCAACTACT 907323
||| : : : : :
317 LeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGln 336
||| : : : : :
907324 TTAGATGAATATCGTCAATATATCGAAAGAGATGCGCACTTTGAACGCGCTTTCCAAAA 907383
||| : : : : :
337 ValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlu 356
||| : : : : :
907384 GTCCTTGTGGACGACCAAGGTAGAGATACCATTCGCGATCTTACGGTGGTTGAAGAA 907443
||| : : : : :
357 LysTyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGln 376
||| : : : : :
907444 CATTATGAATTCATCATCATCATGATATTACTGACCCAGCAATTTGTCGCTGCACACG 907503
||| : : : : :
377 LeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAsp 396
||| : : : : :
907504 CTTTCACATCGTTATATTTCCGATCGTCACTTACCAGATAAAGCCATTGATTGATCGAT 907563
||| : : : : :
397 GluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeu 416
||| : : : : :
907564 GAACGACGCTCTAGCATTCGTATGGAATAATAGATTCTAAACCTGAACCGCTTGATCGCTTT 907623
||| : : : : :
417 GluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLys 436
||| : : : : :
907624 GAACGCTGATTTATCCATTTAAATTTGGAACAACAAGCGTTACAAAAGAGAGAGAGAA 907683
||| : : : : :
437 AlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeu 456
||| : : : : :
907684 GCAAGTCGCAACCGTTTAGAAATGTTAGAAAGAAATTTGGCTGAAAAGAACCGTGAATAC 907743
||| : : : : :
457 GlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeu 476
||| : : : : :
907744 GCGCAACTTGAAGAAAGTATGGAATCTGAAAAGAACGACGCTTTCTCGCTCTCAACATATT 907803
||| : : : : :
477 LysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeu 496
||| : : : : :
907804 AAACAGAGTTAGATCTGCAAAAACCGAACTAGAACAGCTCGTCGCGGGTGATTTA 907863
||| : : : : :
497 AlaArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGln 516
||| : : : : :
907864 GCGAAATGCTCGATGTCATATGCGCGCATCCCTGATCTTGAAGAACCACTTGAGCAA 907923
||| : : : : :
517 LeuGluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHis 536
||| : : : : :
907924 GCTGAACACGCAAGGAAAGAAATGACGCTTTTACGCTACGCTCACAGATGAAGAA 907983
||| : : : : :
537 IleAlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGlu 556
||| : : : : :
907984 ATGCGACAGTGTCTTTAAAGCCACAGCAATCTCTGTATCAAAATGATGGAAGCGGAG 908043
||| : : : : :
557 LysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGln 576
||| : : : : :
908044 AAAGAAAACCTCTGGCTATGGAAGATGAACACTACATAAACGAGTGTGTCAGAAAGAA 908103
||| : : : : :
577 AlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGln 596
||| : : : : :
908104 GCGGTTGATCGGTAGCAACGCGATTCGTGTAGTGTGTCAGGCTCTTCCGATCCCTAAT 908163
||| : : : : :
597 GlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAla 616
||| : : : : :
908164 CGCCCAATTCGTTCTTCTTGTTCCTTAGGCCCAACAGGTGTGCGGAAAACAGAGCTTTCG 908223
||| : : : : :
617 LysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSer 636
||| : : : : :
908224 AAAAATTCGCTAAATTTCTGTTTGTATGATGAAGATGCGTGTGCTATGATATGTCA 908283
||| : : : : :
637 GluTyrMetGluGlnHisSerValSerArgLeuIleGlyValaProGlyTyrValGly 656
||| : : : : :
908284 GAGTTTATGGAACACACAGTGTCTCTGTTAGTTGGTGGCTCCAGCTATGTGGC 908343
||| : : : : :
657 HisGluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeu 676
||| : : : : :
908344 TATGAAGAGCGGTTATTATTAACCTGAAGCTGTTCGTGCTGCTCATATTCAAGTATCTTA 908403
||| : : : : :
677 PheAspGluValGluLysAlaHisValAlaValPheAsnThrIleLeuValLeuAsp 696
||| : : : : :
908404 TTAGATGAAGTTGAAAAGACACACGAGATGATTTCAATATCTTATTAACAGTGTGGAT 908463
||| : : : : :
697 AspGlyArgLeuThrAspGlyGlnGlyArgThrValaAspPheArgAsnSerValIleIle 716
||| : : : : :
908464 GATGGTCTTTAACTGATGTCAAGTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 908523
||| : : : : :
717 MetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMet 736
||| : : : : :
908524 ATGACCTTAACCTGGTCTCGAT---TTAATCCAAGGTAAATAAGACCAAGCTATAGC 908580
||| : : : : :
737 GluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeu 756
||| : : : : :
908580

907264 GGTAATTTGTTAAACCAAGTTTGGCAGGCGAAATTACATTGCGTGGTCAACTACT 907323
 317 LeuGluGluThrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGln 336
 907324 TTAGTGAATATCGTCAATATATCGAATAAGATGCCGCACTTGAACCGCGTTTCCAAAA 907383
 337 ValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlu 356
 907384 GTCTTTGGACGACCAACCAAGTGTAGAGATACCATTCGATCTTACGTGGTTGAAGAA 907443
 357 LysTyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAenAlaGln 376
 907444 CGTATGAATTCATCATCATCGTGGATTAATGACCCAGCAATTTGCGTCGACGACG 907503
 377 LeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAsp 396
 907504 CTTCACATCGTATATTCGATTCGATCGATGATTAACAGCAATTTGATTTGATCGAT 907563
 397 GluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeu 416
 907564 GAACGACGCTAGCATTCGTATGTAATAGATCTTAACCTGACCGCTTGATCGCTTT 907623
 417 GluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLys 436
 907624 GAACGTCGTATATTCCAATTAATAATTGGAACAACAGCGTTTCAAAAAAGAAAGACGAA 907683
 437 AlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspLeuArgAspLysLeu 456
 907684 GCAAGTCGCAACGTTTAGAAATGTAGAGAAAGATTTGGCTGAAAAAGAACCTGGAATAC 907743
 457 GlnProLeuThrMetLysTyrArgLysGluArgIleAspGluIleArgArgLeu 476
 907744 GCCGAACCTTGAAGAAGTATGAAATCTGAAAGACGCGTTCTTGGCTCTCAACATAT 907803
 477 LysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeu 496
 907804 AAACAAGAGTTAGTACTGCAAAACCCGAACTAGAACAGCTCGTCGCGGGTGATTTA 907863
 497 AlaArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGln 516
 907864 CGCAATATGTCAGTTGCAATATGCGCGCATCCCTGATCTTGAAAGCAACTTTGAGCAA 907923
 517 LeuGluGlyThrSerSerGluAsnValMetLeuThrGluAsnValGlyProGluHis 536
 907924 GCTGAACACGACGAAAGAAATGATCAGCTTTACGCTATCGCTCACAGTGAAGAA 907983
 537 IleAlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGlu 556
 907984 ATCGCAGAAGTGTCTTAAAGCCACAGGCAATTCCTGTATCAAAAAATGATGAAGCGAG 908043
 557 LysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlu 576
 908044 AAAGAAAACTTTGCGTATGGAAGATGAATCAATAACAGAGTGAATTTGTCGAAGAA 908103
 577 AlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGln 596
 908104 GCGGTTGATCGGTAGCAACGCGATTCGTCGTAGTCGTGAGGCTCTTCCGATCTCTAT 908163
 597 GlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAla 616
 908164 CGCCCAATGCTTTCTTCTTGTTCCTAGGCGCAACAGGTTTGGGAAAAACAGAGCTTTGC 908223
 617 LysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSer 636
 908224 AAACTTTGGCTAAATCTTTGTTGATGTAAGATGCGGATGCGTATGATATGATGTC 908283
 637 GluTyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGly 656
 908284 GAGTTTATGAAAAACACACAGTGTCTCGTTTGTAGTGGCTCCAGGCTATGTCGC 908343
 657 HisGluGluGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeu 676
 908344 TATGAAGAGCGGCTTATTAACTGAAGCTGTTCGTGCTGCTCAATATTCAGTGATCTTA 908403

QY 677 PheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAsp 696
 Db 908404 TTAGATGAAGTTGAAAAAGCACACGACAGATGATTCATATCAATATCAAGTGTGGAT 908463
 QY 697 AspGlyValGluThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIle 716
 Db 908464 GATGTCGTTTAACTGATGTCGTCAGGTGCTACTGTGGACTCCGTAACACTGTGGTTAT 908523
 QY 717 MetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMet 736
 Db 908524 ATGACCTCTAACTTGGTTCGTGAT--TTAATCCAAGGTAAATAAAGACGAAAGCTATAGC 908580
 QY 737 GluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeu 756
 Db 908581 GAAATG--AAAGCCCTTATGATGTGCTGTGTAAGCCCAACATTTCCGCCCAAGATTCATC 908637
 QY 757 AsnArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysVal 776
 Db 908638 AACCGTATGACGAACCGTGGTATTCATCCATCTTGTAAGAAATATCCGTGCGGAT 908697
 QY 777 AlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAla 796
 Db 908698 GCAAGTATCCATTAAGAACGCTTAGCAAAACGATGGAACCTCGTGGATACGAATTGGTG 908757
 QY 797 ValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAla 816
 Db 908758 TTTACCGATGCTTTTATTAAGCTTCATTTGGCGAAGTGGATACGACCAATTTATGGTGA 908817
 QY 817 ArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValVal 836
 Db 908818 CGTCATTTGAACGCTGCTATTTCACACAGAAATTTGAAACACAGCTTGGCACAAATTTCTA 908877
 QY 837 ArgGluGluIleAspGluAsnSerThrValTyrIleAsp 849
 Db 908878 TCTGTCGCTTATTACCTGCTAAGGTGTTGTTACCATTTGAT 908916

RESULT 6

US-09-103-840A-2/c
 ; Sequence 2, Application US/09103840A
 ; Patent No. 6294328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.
 ; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ; TITLE OF INVENTION: TUBERCULOSIS
 ; FILE REFERENCE: 24366-20007.00
 ; CURRENT APPLICATION NUMBER: US/09/103.840A
 ; CURRENT FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 4403765
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; FEATURE:
 ; OTHER INFORMATION: CDC 1551
 ; OTHER INFORMATION: "n" bases at various positions throughout the sequence
 ; OTHER INFORMATION: represent a, t, c or g
 US-09-103-840A-2

Alignment Scores:
 Pred. No.: 6.03e-205 Length: 4403765
 Score: 2158.50 Matches: 447
 Percent Similarity: 69.66% Conservative: 150
 Best Local Similarity: 52.16% Mismatches: 227
 Query Match: 47.44% Indels: 33
 Gaps: 9

US-09-812-350-17 (1-911) x US-09-103-840A-2 (1-4403765)

2 AsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAla 21
462071 AACCCG-----ACGACCAAGAGCAGCGCGCTTAACCGCGGCTTACAGCGGCT 462021
22 ValAsnAlaGlyHisAlaGlnPheThrProLeuHisIleuAlaGlyAlaLeuIleSerAsp 41
462020 TCGACCGCGCGCAATCCGAGATCCGCGCCTCACCTGTGTATGGGCTGTGACCCAA 461961
42 ProThrGlyIlePheProGlnAlaIleSerAlaGlyGlyGluAsnAlaAlaGlnSer 61
461960 AACGAGGTATGCGCGCACCGCTACTGAGGCTGTGCTGTCGAGCGCGCACCGTCCGC 461901
62 AlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProAsp 81
461900 GCGCAA-----ACCCAGCGCTGCTGACCGCTTTCGCGAGCGGCTGAGCGCACG 461847
82 AspileProAlaSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGlnLys 101
461846 CAGCGCGAGCTGTCCCGAGTGTGTAGCGCGATCACCGCGCGAGCTGCGCCACC 461787
102 SerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGlu---Asp 120
461786 GAGCTGGACGAGAGTACGTCTCCACCGAGCAGCTGATGTGCGGCTGGCCACCGGTGAC 461727
121 SerGlnIleArgAspLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
461726 TCCGAGCTGCGCAAGCTTTGACCGCGCCACCGCGCTCGCGCGAGCGCTGCGGAGGCG 461667
141 ValGluLysLeuArgGlyLysGluGlyLysValGlnSerAlaSerGlyAspThrAsn 160
461666 TTGCTCAAGTGGCGCG-----AGCGCGCGGTACACCGCGCGAACCGGAGCGGAC 461613
161 PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLysLeu 178
461612 TATCAGCGCTGCAGAGTACTTCCACCGACCTGACCGCGCGCGCGCGGAGCAACTC 461553
179 AspProValIleGlyArgAspGluGluIleArgValValArgIleLeuSerArgArg 198
461552 GACCGGTTCATCGCGCGCGCAACAGAGATCCGCGCGGTGTGCGAGGTGTGTCGCGTGC 461493
199 ThrLysAsnProValLeuIleGlyLysProGlyValGlyLysThrAlaValAlaGlu 218
461492 ACCAAGAACACCGGTGCTGATCGGTGAGCGCGCGGTGCGCAAGACCGCGATCGTGAG 461433
219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
461432 GCGCTGGCGCAGCGCATCGTGGCGCGACGTGCGCGAGAGCTTGGCGCAAGACCATC 461373
239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGlu 258
461372 GTCGCGCTCGATCTCGGCTCGATGTGCGCGCTCCCAATACCGCGCGAATTCGAGGA 461313
259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
461312 CGGCTCAAGCGCTCTCTCGACACATCAAGAACTCGCGCGCGCAATCATCATCTTATC 461253
279 AspGlnIleHisLeuValLeuGlyAlaGlyLysThr---GluGlySerMetAspAlaAla 297
461252 GACGAGCTGCACACCATCGTGGCGCGCGCGCCACCGCGAGGGGCGGATGCGCGCGC 461193
298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeu 317
461192 AACATGATCAAGCGATCTGCGCGCGCGCGGAGTTACGGCTGGTGGCGGCGACACCGCTG 461133
318 GluGlnTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnVal 337
461132 GACGAATACCGCAAGCACATCGAGAGAGACCGCGCGCTCGAGCGCGGTTCACACAGGTG 461073
338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLys 357
461072 TACGTGCGGAGCGCGTGGTGGAGACACCATCGCATCTCTGCGCGCGCTCAAGACCGC 461013
358 TyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377

Db 461012 TACGAGGTGCACCCAGCGGTGCGCATCCGACTCGCGCTGTCGACGCTGCTTGG 460953
QY 378 SerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
Db 460952 AGCGACCGGTATATACACCGCGCTTCTTGGCGCGCAAGGCCATCGACCTGTGTCAGCAG 460893
QY 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGlu 417
Db 460892 GCGGCCAGCGCGGTGGATGAGATCGACTCGCGCGCTGTCGATCGACGAGTCCAG 460833
QY 418 ArgLysArgMetGlnLeuGluIleGlnLeuHisAlaLeuGluArgGluLysAspLysAla 437
Db 460832 CGGCTGTCGCGCGGTGGAGATCGAAGAGATGCGCTGTCCAAAGAGACGAGCGCG 460773
QY 438 SerLysAlaArgLeuIleLeuValArgLysGluLeuAspAspLeuArgAspLysLeuGln 457
Db 460772 TCGCGCGAGCGGTGGCAAGCTGCGCTCGGAGCTGGCGGACCCAGAAAGAGAGTTGGCC 460713
QY 458 ProLeuThrMetLysTyrArgLysGluLysGluA---GileAspGluIleArgArgLeuLys 477
Db 460712 GAGCTCACCCGCTGGCAGAACCGAAGAACGCGATCGAAATCGTCCGCGACCTCAAG 460653
QY 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAla 497
Db 460652 GAGCAGCTGGAAGCCCTGCGCGGGAATCCGAGCGCGCGCAACGCGACCGCTGCGC 460593
QY 498 ArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIle---AlaGln 516
Db 460592 AAGCGCGCGAGCTGCGCTACGAGCGATCCCGAGGTGGAGAAAGCTCGACCGCGCG 460533
QY 517 LeuGluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHis 536
Db 460532 TTGCGCGACGCGCAGCGCGGAGCAGGTGATGCTCAAGGAGAGTCTCGTCCGACGAC 460473
QY 537 IleAlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGlu 556
Db 460472 ATCGCGCGAGTGTGTCGCGGTGGAACCGGCAATCCCGCGCGCTCGCTCGAAGCGCG 460413
QY 557 LysGluArgLeuIleLysLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGln 576
Db 460412 ACCGCCAAGCTCTGCGCATGGAAGACGAGCTGGCGACCGGTCTATCGGCGAGAGGCC 460353
QY 577 AlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGln 596
Db 460352 GCGGTTACCGCAGTCTCTGATCGGTGCGCGCACCGCGCGCGGTGTCCGACCCCAAC 460293
QY 597 GlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAla 616
Db 460292 CGGCCCCACCGCGCGGTTCATGTTCTCGCGCGCAACCGGTGTGCGCAAGACCGAGTGGCC 460233
QY 617 LysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSer 636
Db 460232 AAGCGCTGCGCGCATCTCTGTTTCGACGACGAGCGGCGATGGTCCGATCGCATGAGC 460173
QY 637 GluTyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGly 656
Db 460172 GAGTACCGCGAGAACACACCGTGTGTTGATCGCGCGCGCGCGCGCTATGTGGA 460113
QY 657 HisGluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeu 676
Db 460112 TACGAGCGCGCGGTGAGTGTGCGCGCGGTGCGCGCGGTTCMTACCGGTGGTGTG 460053
QY 677 PheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAsp 696
Db 460052 TTCGACGAGATCGAGAGAGCGCACCGCGACGTGTTCGAGGTGCTGTCAGGTCTCTCGAC 459993
QY 697 AspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIle 716
Db 459992 GAGGCGCGCTCACCGACGCGGACCGCGCGCTGCTGCGCTTCCGCAACACCATCTTGTATC 459933
QY 717 MetThrSerAsnLeuGly-----AlaGluHisLeuLeuAlaGlyLeuThrGly 732

459932 CTGACGTCCAAACCTGGGTGGGTGGCAGCGCGAGCGAGTGTGGTCCCGCGG----- 459882
 733 LysValThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArg 752
 459881 -----GTGGCGCTACGTTTCAAG 459864
 753 ProGluLeuLeuAsnArgLeuAspGluIleValValPheAspProLeuSerHisAspGln 772
 459863 CCGAGTTTATCAACCGGCTCGACGAGTGTCTATCTTTGAGGGTCTCAACCCCGAAGAG 459804
 773 LeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGly 792
 459803 CTGGTGGCATCTCGACATCTCAGTGGCGGCGAGCGCAAGCGGTGGCGAGCGCGG 459744
 793 ValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspPro 812
 459743 CTCACGTCTGAGGTCTCTGTCGCGGCAAGCGTGTGGCGAGCGCGGATTTCGACCCG 459684
 813 ValTyrGlyValAlaArgProIleArgAlaGlyTyrMetGluLysLysValValThrGluLeuSer 832
 459683 GTGTACGGGGCGGCGGTGTGCGCGGTGGTGCAGCAGCCATCGTGCACGAGTGGCC 459624
 833 LysMetValValArgGluGluIleAspGluAsnSerThrValTyrIleAsp 849
 459623 AAGATGCTGTGGCGCGGCGAGGTGCAGCAGCGGATACCGTGGCGGTCAAC 459573
 LT 7
 9-103-840A-1/c
 quence 1, Application US/09103840A
 tent No. 6294328
 NERAL INFORMATION:
 PPLICANT: FLEISCHMAN, Robert D.
 PPLICANT: WHITE, Owen R.
 PPLICANT: FRASER, Claire M.
 PPLICANT: VENTER, John C.
 TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ILE OF INVENTION: TUBERCULOSIS
 ILE REFERENCE: 24366-20007.00
 URENT APPLICATION NUMBER: US/09/103,840A
 URENT FILING DATE: 1998-06-24
 UMBER OF SEQ ID NOS: 2
 OPTWARE: PatentIn Ver. 2.1
 Q ID NO 1
 LENGTH: 4411529
 TYPE: DNA
 ORGANISM: Mycobacterium tuberculosis
 OTHER INFORMATION: H37Rv
 9-103-840A-1
 nment Scores:
 . No.: Length: 4411529
 e.: 2158.50 Matches: 447
 ent Similarity: 69.66% Conservative: 150
 Local Similarity: 52.16% Mismatches: 227
 Y Match: 47.44% Indels: 33
 3 Gaps: 9
 9-812-350-17 (1-911) x US-09-103-840A-1 (1-4411529)
 2 AsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAla 21
 461988 AACCCG-----ACGACCAAGACGAGCGCGCTAACCGCGGTACAGCGGCT 461938
 22 ValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSerAsp 41
 461937 TCACCGCGCGCAATCCCGAGATCCGCGCGCTCACCTCTGATGCGGTGCTACCCAA 461878
 42 ProThrGlyIlePheProGlnAlaIleSerAlaGlyGluAsnAlaGlnSer 61
 461877 AACGACGATTCGCGCACCGTACTGAGGCTGTGGGTGTGAGCCCGCACCGTCCGC 461818
 62 AlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProAsp 81
 460749 CGGTGTGGTGGCGGTGGGATCGAAGAGATCGCGTGTCCAAAGACGAGGCGG 460690
 461817 GCCGAA-----ACCCAGCGCTGTCTGACACCGTTCGCGAGCGAGCTGGAGCCAGCACG 461764
 82 AspileProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGlnLys 101
 461763 CAGCGCAGCTGTCCCGAGTCTGTAGCGCGGATCACCACCGCGCAGCTGGCCACC 461704
 102 SerArgGlyAspThrHisLeuAlaValAspGlnLeuMetGlyLeuLeuGlu---Asp 120
 461703 GAGCTGGACGACGAGTAGTCTCCACCGAGCACGTGTGTGTGGTGGCGGCTGGCCACCGGTGAC 461644
 121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
 461643 TCCGACGTGCGCAAGCTGTTGACCGGCCACGCGCCCTCGCGAGGCGCTCGCGGAGCGG 461584
 141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
 461583 TTCGTCAAGGTGCGCGG-----AGCGCGCGGTCTACCGCCCGGAAACCGAGCGCGACC 461530
 161 PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLysLeu 178
 461529 TATCAGGCGCTGCAGAAAGTACTCCACCGACCTGACCGCGCGCGCGCGGCGGCAAACTC 461470
 179 AspProValIleGlyArgAspGluIleArgValValArgIleLeuSerArgArg 198
 461469 GACCGGTCTATCGCGCGGCAACAGAGATCGCGCGCTGTGTGAGGTGTCTCCGTCGC 461410
 199 ThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGlu 218
 461409 ACCAAGAACACCGCGTGTGATCGGTGAGCGCGGTCTCGGCAAGACCGCGATCGTGAG 461350
 219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
 461349 GCCTGGCGCAGCGCATCTGTGGCGCGGCGAGCTTGGCGGAGAGCTTGGCGCAACACCATC 461290
 239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGlu 258
 461289 GTCCGCTCGATCTCGCTCGATGTGTGCGGTCTCCAAATACCGCGCGGATTCGAGGAA 461230
 259 ArgLeuLysSerValLeuLysGlyValGluAspAlaGlyLysValIleLeuPheIle 278
 461229 CGGCTCAAGCGCGTCTCGACGACATCAAGAACTCGCGCGGCCCAATCATCATCGTTTCATC 461170
 279 AspGluIleHisLeuValLeuGlyAlaGlyLysThr---GluGlySerMetAspAlaAla 297
 461169 GACGAGTGCACACCATCTGTGGCGCGCGCGCGCACCGCGCGCGGCGATGGACGCGCGG 461110
 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeu 317
 461109 AACATGATCAAGCGGATGTGGCGCGCGGCGAGTTACGGTGTGTGGGGCGCACCGCTG 461050
 318 GluGluTyrArgLysTyrValGluLysAspAlaPheGluArgArgPheGlnGlnVal 337
 461049 GACGATACCGCAAGCACATCGAGAGAGCGCGCGCTCGAGCGCGCTTTCACACAGGTG 460990
 338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLysGluLys 357
 460989 TAGCTCGCGAGCGCTGCGGTGGAGACCATCGGATCTCTGCGCGGCTCAAGACCGG 460930
 358 TyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377
 460929 TAGGAGTGCACACCGGCTGCGCATCACCAGATCTCGCGGTGTGTGGCAGCTGCCATTG 460870
 378 SerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 460869 AGCGACCGGTATATCAACCGCGCTCTCTCGCGCACAGGCCATCGACCTGTGTGACGAG 460810
 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluIleAspAsnLeuGlu 417
 460809 CGCGCCACCGCTGCGGTGGAGATCGACTCGCGCGCTCGAGATCGACGAGGTGCGAG 460750
 418 ArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAla 437
 460749 CGGTGTGGTGGCGGTGGGATCGAAGAGATCGCGTGTCCAAAGACGAGGCGG 460690

438 SerLysAlaAArgLeuLeuValArgLysGluLeuAspLeuArgAspLysLeuGln 457
 460689 TCGGCGAGCGGTGGCCAAAGTCCGCTCCGAGCTGGCGACCAAGAGAGAGTGGCC 460630
 458 ProLeuThrMetLysTyrArgLysGluLysGluArgHisLeuAspGluLeuArgLysLeu 477
 460629 GAGCTCACACCCCGTGGCAGAACCGAGAGAACCGCATCGAATCTTCGCGAGCTCAAG 460570
 478 GlnLysArgGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAla 497
 460569 GAGAGCTGGAAGCCCTCGCGGGGGAATCCAGCGCGGCGAGACCGCGAGCTGGCC 460510
 498 ArgAlaAlaAspLeuArgTyrGlyAlaLeuGlnGluValGluSerAlaLeu 516
 460509 AAGGCGCGCGAGCTCGCTACGAGCGCATCCCGAGTGGAGAGAGCTCGACGCGCG 460450
 517 LeuGluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHis 536
 460449 TTGCGCAGCGCAGCGCGGAGCAGGTGATGCTCAAGGAGGAGGTGCTCCCGCAGC 460390
 537 IleAlaGluValValSerArgTyrThrGlyLeuProValThrArgLeuGlyGlnAsnGlu 556
 460389 ATCCGCGAGCTGTGCTCGCGTGGACCGGCGATCCCGCGGCTCGGCTGCTGGAAGCGAG 460330
 557 LysGluArgLeuGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGln 576
 460329 ACCGCCAAGCTGCTCGCATGAGAACGAGCTGGCAAGCGGTGTCATCGGCGAGAGGCC 460270
 577 AlaValAsnAlaValSerGluAlaLeuLeuArgSerArgAlaGlyLeuGlyArgAlaGln 596
 460269 GCGGTACCGCAGTCTCTGATGCGTGGCGGCGACCGCGCGGCTGGCGACCCCAAC 460210
 537 GlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAla 616
 460209 CGGCCACCGCGCGGTTCATGTTCTTCGCGCGCGACCGGTGTCGCGCAAGCGAGCTGGCC 460150
 617 LysAlaLeuAlaGluGlnLeuPheAspAspGluLeuValArgLysLeuMetSer 636
 460149 AAGCGCTGGCGGACTTCTGTTGACGACGAGCGGCGGATGGTCCGATCGATGAGC 460090
 637 GluTyrMetGluGlnHisSerValSerArgLeuLeuGlyAlaProGlyTyrValGly 656
 460089 GAGTACGCGAGAGACACACCGTGGCTCGGTGATCGCGCGCGCGCGCTGATGTGGGA 460030
 657 HisGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValLeuLeu 676
 460029 TACGAGCGCGCGGTTCAGCTGACCGAGCGGTGGCGCGCTCCCTACCGTGGTGTG 459970
 677 PheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAsp 696
 459969 TTCGACGAGATCGAAGAGCGCACCCCGACGTGTCGACGCTGCTGCTGCGAGGTCTCGAC 459910
 697 AspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValLeuLeu 716
 459909 GAGGCGCGCTCACCGACGGCACCGCGCGACGTGCTGCTTCGCGACACCATCTTCATC 459850
 717 MetThrSerAsnLeuGly-----AlaGluHisLeuLeuAlaGlyLeuThrGly 732
 459849 CTGACGTCCAACTGGGGTGGGTGGCGAGCGCGAGCGGTGCTGGCGCG----- 459799
 733 LysValThrMetGluValAlaArgAspCysValMetArgGluValArgLysPheArg 752
 459798 -----GTGCGCGCTTACGTTCAAG 459781
 753 ProGluLeuLeuAsnArgLeuAspGluLeuValValPheAspProLeuSerHisAspGln 772
 459780 CCGAGTTCATCAACCGCTCGACGACGTGCTCATCTTTGAGGTCTCAACCCCGAAGAG 459721
 773 LeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgLys 792
 459720 CTGGTGGCATCGTTCGATCATCGTGGCGAGCTGGCGAAGCGGTGGCGAGCGCGG 459661

793 ValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspPro 812
 459660 CTCAGCTGCAGGTCTCGCTGCGCGCAAGCGTGTGGCGAGCGCGGATTCGACCGG 459601
 813 ValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSer 832
 459600 GTGTACGGGCGCGCGCTGTGGCGCGCTGGTGGCAGCAGCGCATCGTGCACGCTGGCC 459541
 833 LysMetValValArgGluGluLeuLeuAspGluAsnSerThrValTyrIleAsp 849
 459540 AAGATGCTGTGGCGCGCGCGGAGTGGCAGCGCGGATACCGTGGCGGTCAAC 459490

RESULT 8
 US-09-328-352-2604
 ; Sequence 2604, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 2604
 ; TYPE: DNA
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-2604

Alignment Scores:
 Pred. No.: 2,96e-208 Length: 2592
 Score: 2140.00 Matches: 424
 Percent Similarity: 70.35% Conservative: 181
 Best Local Similarity: 49.30% Mismatches: 245
 Query Match: 47.03% Indels: 10
 Gaps: 6

US-09-812-350-17 (1-911) x US-09-328-352-2604 (1-2592)

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 Db 13 ATGCGATTGTGAATAATTACGAACCGCTTGCAGCAGCCCTCTCAGATGCTCAATCCTTA 72
 QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuSer 40
 Db 73 GCGATGGTAAAGACCACCATACAGCTATAGCAGGTATTCATATTTGAGTACTTTATTGAA 132
 QY 41 AspProThrGlyLeuPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
 Db 133 GAGCGCTCCAATATT-----AGTTTGTTCACCAAGCAGGTGCGACGG-----TTACT 180
 QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProPro 80
 Db 181 GAACCTTAAACAAAGCTAGACGAGCTTTAAAGATGCTCCCACTATTGCTAACCCGAGC 240
 QY 81 AspAspIleProAlaSerSerSerLeuLeuLysValIleArgArgAlaGlnAlaAlaGln 100
 Db 241 GCGCATGTCAATTTAAACCCAGCAGTTAAAGCACTCAACTTGGCAGATCGATACGCG 300
 QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuAsp 120
 Db 301 CAAAAGCTGGCGATGAGTGTGTTGCTCACTGCTGGGTGTTTATTGGGCTTGGCGAACT 360
 QY 121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaIleThrAlaArgValLysSerGlu 140
 Db 361 GGTGAACCAAAATATTTAAGTCCGCTAGTGTGTAATCTCCGACGAGCTTACGCAAGTA 420
 QY 141 ValGluLysLeuArgGlyLysGlyLysValGluSerAlaSerGlyAspThrAsn 160
 Db 421 ATTGAATAATTCGAGGTAGTGAC-----AAAGTCATGAGTAATAATACGAGACCAA 474
 QY 161 PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLysLeu 178


```

475 CGTGACTACCTTAATAATATACGATTGATTAAACCGCGGCTTTATCGGGAACCTT 534
179 AspProValIleGlyArgAspGluGluIleArgValValArgIleLeuSerArgArg 198
535 GATCCAGTGATTGGCGCGTGATGAGATCGCGGTACCATTCAGTCTTGTACGCGGT 594
199 ThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGlu 218
595 ACTAAAATAACCCAGTACTATTGGTGAACCTCGGGTAGGTAAACCGCTATTGTTGAA 654
219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
655 GGTTCGCACAACTGATTGTCAATGGTGAAGTACCAAGAGCTTAAAGATAAACGTT 714
239 IleSerLeuAspMetGlyAlaLeuValAlaGlyLysThrArgGlyGluPheGluGlu 258
715 TTATCGTTAGATTAGGTTTCAATCTTGCAGTCTCAAGTATCGTGTGAGTTTGAAGAG 774
259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
775 CGTTAAAGCTGTTTAAAGATTGGCGAAACACCAAGAGCGAAATCATCTTATTCATT 834
279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsn 298
835 GACGAGTTACATACACTCGTTGGTGGTAAAGGTGACGCGCGATGGATGCAGGTAAT 894
299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
895 ATGTTAAACACCTGGTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 954
319 GluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnInValTyr 338
955 GAATATCGCCATACATTGAAGAATGACGCTTGGAGCGTCTTTCCAAAGAGTCTG 1014
339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyr 358
1015 GTCGATGAACCAAGTAGAAGATACCATTCGATTGCGATTTCAGTGGTGTGAAAGAAAGTAT 1074
359 GluGlyHisHisGlyValArgIleGluAspArgAlaLeuIleAlaGlnLeuSer 378
1075 GCGACTCACCATGCGGTACGATTTTGAATTCAGCTCAGCATTTATGTCGCGGCAAAATGCT 1134
379 AlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
1135 CATGTTATATTACAGACCGTCAATTACAGACAGCGGATGACCTGATTGATGAGGCC 1194
399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArg 418
1195 GCTTCTCGTATTAAAGATGGAATCGATTCTTAAGCCAGACGACTTGTATAAATTTGATCGC 1254
419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSer 438
1255 GGTTCATCCAGTTGAAATGCAATTCGAACCGGTGAAAGAAAGATGAAGACGCGAGCAGT 1314
439 LysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnPro 458
1315 AAGCCCAAGTTACTCATCTCTGAATAACAGATCGCTGAAGTCGAGAAAGATACACGAT 1374
459 LeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGln 478
1375 CTGAAGAAGTGTGGAAGCTGAGAAAACACTGTGTAGAGGCACTAAACAAAGCTCAGGTT 1434
479 LysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArg 498
1435 GAATCTGAAGACCGTATTGCTTTTGAATAAGCTCAGCGCAAGCGCATTTTGGCAGAA 1494
499 AlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGlu 518
1495 GCACGAGTTTGCATATGATGTTGATTCAGAGTTGCAAAAACAACTTGAACAGATGAA 1554
519 ---GlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537

```

RESULT 9

US-09-199-637A-280

; Sequence 280, Application US/09199637A

; Patent No. 6355411

; GENERAL INFORMATION:

; APPLICANT: Ausubel, Frederick

; APPLICANT: Goodman, Howard M.

```

1555 GTTGTGTAAGAAAAACGAAGAGCCAAACTCATTCGTACAAAAGTAAGTGAATAATGAAATT 1614
538 AlaGluValValSerArgTrpThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
1615 GCCGAAGTCGTTAGTGTCTCAACAGATATTCCGTTGCTTAAATATGATCAAGCGCAGCGT 1674
558 GluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
1675 GAAAAAAGCTCTTCATATGAAGAGATTCTTGCATGACCCGTTGTAGGGCAAGATGAAGCA 1734
578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGln 597
1735 GTTCGTTCCGTTATCGAATGCTGTTCCCGTTCCTGCTGCTGTTCTGACCCGATCGT 1794
598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGlnLeuAlaLys 617
1795 CCTAGCGGATCATCTTCTGTTTGTAGGACCAACAGGTGTGGTAAAACTCAGTTGACTAAA 1854
618 AlaLeuAlaGluLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGlu 637
1855 GCTTTAGCTAACTTCTTGTTCACAGTGAATGATGCCATGATTCGTATCGATATGATGAA 1914
638 TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHis 657
1915 TTTCATGGAGAAACATTCGTACGCGTTTAGTTGGTGGCTCCGCTCCGGTTATGTAGTTAC 1974
658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe 677
1975 GAAGAGGGCGGTGTTTAAACCAAGCTGTTCCCGTAAACCATATAGCGTAGTGTGTT 2034
678 AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAsp 697
2035 GATGAGTTGAAAAAGCGCATCCAGATGCTTCAATATCTTGTACAAAGTGTGACCAT 2094
698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMet 717
2095 GCAAGCTTAAACGACTCAACAGTCTGTAGTGGACTTTAAAAACACCGTTATTGTGATG 2154
718 ThrSerAsnLeuGlyAlaGluHisLeuAlaGlyLeuThrGlyLysValThrMetGlu 737
2155 ACATCGAACTTCGGTTTCAAGATGTCGTAACCTTGGTGAAGGTGCAACTGATGATGAA 2214
738 ValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsn 757
2215 GTG---CGTACTATTGTAATGAATCGGTAAAGTACAGCATTTCCGTCAGAAATTTATTAAAC 2271
758 ArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAla 777
2272 CGGATTGATGAGTGTGATTTTCCATTCACCTCAAAAAGCACAGATTCGTGGCATTGCC 2331
778 ArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaVal 797
2332 GATATTCAGTTGGACCGCTTACGCTCACGACTTGTGTGATCGTGTATGATGATTAAGTCTGA 2391
798 ThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyValArg 817
2392 GATGACAGTGCAATTTGACTTATGATTGACGCTGGTTTGTATCTCTATATACGAGCGCGT 2451
818 ProIleArgArgTrpMetGluLysValValThrGluLeuSerLysMetValValArg 837
2452 CCATTGAAACGTGCAATTCACACAGGTTGAAATAACACTAGCTCAAAAAATCTTGTCA 2511
838 GluGluIleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyr 857
2512 GGTGACTTTTGGCGGTGATACCATTTTAAAGGTGAAATGGTCACTTAGTGT 2571

```

APPLICANT: Rahme, Laurence G.
 APPLICANT: Mahajan-Miklos, Shalina
 APPLICANT: Tan, Man-Wah
 APPLICANT: Cao, Hui
 APPLICANT: Drenkard, Eliana
 APPLICANT: Tsongalis, John
 TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
 FILE REFERENCE: 00786/361002
 CURRENT APPLICATION NUMBER: US/09/199,637A
 CURRENT FILING DATE: 1998-11-25
 PRIOR APPLICATION NUMBER: 60/066,517
 PRIOR FILING DATE: 1997-11-25
 NUMBER OF SEQ ID NOS: 437
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 280
 LENGTH: 2580
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 09-199-637A-280
 Alignment Scores:
 Length: 2580
 Matches: 427
 Mismatches: 176
 Indels: 245
 Gaps: 6
 Percent Similarity: 69.95%
 Local Similarity: 49.54%
 Match: 14
 Gap: 6
 09-812-350-17 (1-911) x US-09-199-637A-280 (1-2580)
 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 16 ATGCGAATAGACCGTTTTCACCAAGAGTCACTGGCGCTCTCCGAGCGCCAGTCCCTG 75
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuSer 40
 76 GCGGTGGCCATGACCATCCGCGCATCCGCGCGTGGACCTGCTTTCGCGCTCTGCTGAG 135
 41 AspProThrGlyPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
 136 CAGCAAGCGGTTTCGATCAAGCCCTGCTGATGACAGTGGTTCGATATCGCGCCCTG 195
 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProPro 80
 196 CGCAGCGG-----CTCAACAAGAACTCGACCGCGCTGCCGAAGATCCAGAGCCGAC 249
 81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGln 100
 250 GCGCAGCGTAACCTGCCAGGATCTCGACGCGCTGCTCAACAGCTGACCGCTGCGCC 309
 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuLeuMetGlyLeuLeu---Glu 119
 310 CAGCAGAAGGCGGACCACTTCTCCAGCAGAGTGGTATTGCTGGCGCGATGACGAG 369
 120 AspSerGlnIleArgAspLeuLeuAsnGlnValGlyValAlaThrAlaArgValLysSer 139
 370 AACACCGCTCGGCACTGCTGCTCGCGCAGGCGGTGTCGCGCAAGCGCGCTCGAAGT 429
 140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159
 430 GCGGTGGCAACCTGCGTGGC-----GGCAAGCGGTGAACGACCGCCGACGTCGAGAG 483
 160 AsnPheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAla-----GlyLys 177
 484 TCGCGCCAGGCGCTGGACAAGTACACCGCTGCATGACCAAGCGCGCGCGGAAGGCAAG 543
 178 LeuAspProValIleGlyArgAspGluGluIleArgArgValValAlaGlyLeuSerArg 197
 544 CTCGACCGGTGATCGGTGGCAGCAGATCCCGCGACCATCCAGGTCCTGACGCG 603
 198 ArgThrLysAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217

Db 604 CGGACCAAGAAACCAACCGGTGCTGATCGGCGAAACCCGCGCTCGCAAGACCGCCATCGTC 663
 Qy 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 Db 664 GAGGCTGCGCCAGCGCATCATCAAGCGCAAGTCCGCGAGCGCTCAAGGACAAGCGC 723
 Qy 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrglyGlyGluPheGlu 257
 Db 724 CTGCTGCGCCCTGGACATGGGGCGCTGATCGCGGTGCCAAGTTCGCGCGAGTTCGAG 783
 Qy 258 GluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPhe 277
 Db 784 GAACGCTGAAGCGGTCTCTCAACGAACCTGGCAAGCAGCAAGCGCGGTCTATCTCTGTT 843
 Qy 278 IleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAla 297
 Db 844 ATCGACGACTGCACCATGTTGCGCGCGCGCAAGCGGAGGTGCCATGGACCGCGCG 903
 Qy 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrLeu 317
 Db 904 AACATGCTCAAGCGGTCTGCGCGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 963
 Qy 318 GluGluTyrglyLysTyrglyValGluLysAspAlaAlaPheGluArgPheGlnVal 337
 Db 964 GACGAGTATCGCCAGTACATCGAAGAGTTCGCGCTGCGAGCGCGCTTCCAGAGGTG 1023
 Qy 338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLysGluLys 357
 Db 1024 CTGCTGGACGACCGAGCGAGGAGACCATCGCTCTCTGCTGCTGCTGCTGCTGCTGCT 1083
 Qy 358 TyrGluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377
 Db 1084 TATGAAGTGCACCGCGGTGAGCATCCAGCGCGCGATCATCGCGCGCGCGCGCGCGCTG 1143
 Qy 378 SerAlaArgTyrglyThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 Db 1144 TCACACGCTTACATCAACCGATCGGCAACTGCGCGACAGCGCATCGATCGATCGACGAG 1203
 Qy 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGlu 417
 Db 1204 GCGCGCAGCGCATCGCATGAGTCTGACTCAAGCGCGAGGAACTGATCGTCTGAC 1263
 Qy 418 ArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAla 437
 Db 1264 CPTCGCTGATTCAGCTGAAGTTCGAGCGCGAGCGCTGAGAGAGAGAGAGAGAGAGAG 1323
 Qy 438 SerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGln 457
 Db 1324 ACCAGAAAGCGCTGCGCAAGCTGGAGGAGGATATCGTCAAGCTCGAGCGCGAATACGCC 1383
 Qy 458 ProLeuThrMetLysTyrglyLysGluLysGluArgIleAspGluIleArgArgLeuLys 477
 Db 1384 GACCTCGAGGAGATCTGGAAGTCCGAGAGCGCGAGGTGCGAGGCTCGCGCGCATCCAG 1443
 Qy 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrglyAspLeuAla 497
 Db 1444 CAGAAGATCGACGAGCCCAAGCAGGAGATGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 1503
 Qy 498 ArgAlaAlaAspLeuArgTyrglyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 517
 Db 1504 AGCATGCGCGCATCCAGTACAGACCATCCCGACCTCGGACCTGGAAACGCGCTCGAGATGTC 1563
 Qy 518 GluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
 Db 1564 GACCAAGCGCGCAAGCAGAGACCGAGTCTGTCGCAACAGGTGACCGCAGCAGGAAATC 1623
 Qy 538 AlaGluValValSerArgTyrglyThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
 Db 1624 GCGAAGTGTGTTCCAAAGTGGACCGGTATCCCGGTGTCGAAGATGCTCGAGCGCGCGCG 1683
 Qy 558 GluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
 Db 1684 GAGAAGCTGCTGCGCATGAGCAGGAGGCTGCTGCGCGAGTGTGCTGCGCGCGAGGCGG 1743

```

578 ValAenAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyValArgAlaGlnGln 597
1744 GTAGTCGGCGTGTCCACACGGCTGGCGGCTGGCGGCGGCTGGCGGCTGGCGGCTGGCGG 1803
598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLys 617
1804 CGAGCGGCTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1863
618 AlaLeuAlaGluGlnLeuPheAspGluLeuLeuValAlaGlyLeuValAlaGlyLeu 637
1864 GCGCTGGCGGCTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1923
638 TyrMetGluGlnHisSerValSerArgLeuLeuGlyAlaProProGlyTyrValGlyHis 657
1924 TTCATGAGAAACACATCGGTGGCGGCTGGTGGCGGCTGGTGGCGGCTGGTGGCGGCTTC 1983
658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValLeuLeuPhe 677
1984 GAGGAGCGGCTACCTACCGAGCGGATCCCGCGGATCCCGCGGATCCCGCGGATCCCG 2043
678 AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAsp 697
2044 GACGAGGTGGAGAGGCCATCCGAGGTATTCACATTCCTCTCCAGGTGCTCGAGGAC 2103
698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMet 717
2104 GGAGCGCTGACGACAGTACGCGGCTGACGCGGCTGACGCGGCTGACGCGGCTGACG 2163
718 ThrSerLeuGlyAlaGluHisLeu-----LeuAlaGlyLeuThrGlyLysValThr 735
2164 ACCTCCAACTCGGTTCGGCGCAGATCCAGAGCTGGCGCGC-----GAC 2208
736 MetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeu 755
2209 CGCAGGCGCACTGCGCGGCTGATGACGCGGCTGATGACGCGGCTGATGACGCGGCT 2268
756 LeuAsnArgLeuAspGluLeuValValPheAspProLeuSerHisAspGlnLeuArgLys 775
2269 ATCAACCGGATCCAGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2328
776 ValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeu 795
2329 ATCGCGAGATCCAGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTG 2388
796 AlaValThrAspAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGly 815
2389 GAATGAGCGAGGCGCTGGCAAGCTGATGCGGCTGATGCGGCTGATGCGGCTGATGCG 2448
816 AlaArgProLeuArgArgTyrMetGluLysValValThrGluLeuSerLysMetVal 835
2449 GCACGCGCGCTGAAGCGGCGCATCCAGCGCTGGATCGAGAACCGCTGGCGCAACTGATC 2508
836 ValArgGluGluLeuAspGluLeuSerThrValTyrIleAspAlaGlyAlaGlyAspLeu 855
2509 CTGCGCGCAATTCGCGCGGCTGGCGGCTGATCTCGCGGCAAGGTGGAAGCGCGAGATC 2568
856 ValTyr 857
2569 GTCTTC 2574

```

LT 10

09-199-637A-272/c
sequence 272, Application US/09199637A
tent No. 635411

GENERAL INFORMATION:

APPLICANT: Ausubel, Frederick

APPLICANT: Goodman, Howard M.

APPLICANT: Rahme, Laurence G.

APPLICANT: Mahajan-Miklos, Shalina

APPLICANT: Tan, Man-Wah

APPLICANT: Cao, Hui

APPLICANT: Drenkard, Eliana

```

; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 007867/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR FILING DATE: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 272
; LENGTH: 2970
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-272

```

```

Alignment Scores:
Pred. No.: 1,91e-207 Length: 2970
Score: 2133.00 Matches: 440
Percent Similarity: 68.90% Conservative: 178
Best Local Similarity: 49.05% Mismatches: 257
Query Match: 46.88% Indels: 24
DB: 4 Gaps: 8

```

US-09-812-350-17 (1-911) x US-09-199-637A-272 (1-2970)

```

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
Db 2949 ATGCGAATAGACGGTTTGACCAAGCAAGCTGCAACTGGCGGCTCTCCGAGCGCCAGTCCCTG 2890
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
Db 2889 GCGGTGGCCATGACCATCCGCGCATGAGCGGCTGACGCGGCTGACCTCTTCCGCTGCTCGAG 2830
QY 41 AspProThrGlyPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
Db 2829 CAGCAAGGCGGTTCGATCAAGCCCTGCTGATGACGAGTGGCTTCGATATCGCGCCCTG 2770
QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProThr 80
Db 2769 CCGACGCGG-----CTCAACAAGAACTCGACGCGCTGCCGAAGATCCAGAGCCCGACC 2716
QY 81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGln 100
Db 2715 GCGGACGTGAACCTGTCGAGGATCTCCAGCGCTGCTCAACAGGCTGACCGCTGCGCC 2656
QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu---Glu 119
Db 2655 CAGCAGAAGGCGCACCATTCATCTCCAGCGAGCTGTTATGCTGGCGCGCATGACGAG 2596
QY 120 AspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
Db 2595 AACACGAGCTCGGCAAGCTGCTGCTCGGCGGCGGCTGTCGCGCAAGGCGCTGGAGAT 2536
QY 140 GluValGluLysLeuArgGlyLysGluGlyLysValValGluSerAlaSerGlyAspThr 159
Db 2535 GCGGTGGCCAACTCGGTGGC-----GGCAAGCGGTGAACGACCGCGAGGAGGCAAG 2482
QY 160 AsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLys 177
Db 2481 TCGCGCCAGGCGCTGGACAAAGTACACCGTCGACATGACCAAGCGCGCGAGGAGGCAAG 2422
QY 178 LeuAspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArg 197
Db 2421 CTCGACCGCGGTGATCGGTGCGAGCAGAGATCCGCGGACCATCCAGGTCTCTGACGCG 2362
QY 198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217
Db 2361 CGGACCAAGAACACCGCGGTGCTGATCGGCGAACCCCGGCTGCGCAAGACCGCATCGTC 2302
QY 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
Db 2301 GAGGCGCTGGCGCAGCGCATCATCAACGCGGAGTGGCGGACGCGCTCAAGGACACGCGC 2242

```

598	QY	ProThrGlySerPheIeuPheIeuGlyProThrGlyValGlyLysThrGluLeuAlaLys	617
1161	Db	CGAGCGCGCTGTTCTCTCTCCGCGCCGACCGGGTGGGCAAGACCGAGTTGTGCACG	1102
618	QY	AlaLeuAlaGluLeuIleuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGlu	637
1101	Db	GGCTGGCGAGTTCTCTTCGATACCGAGGAGGGCGCTGGTGGCGATAGATATGTCGAG	1042
638	QY	TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHis	657
1041	Db	TTTCATGGAGAAACACTCGGTGGCCCGCTGATCGGCGCCTCCGGGCTACGTCGGCTTC	982
658	QY	GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleIleuPhe	677
981	Db	GAGGAAGCGGCTACTGACCGAGGGGATCCGCGCAAGCCCTACTCGGTGGTGTCTGCTG	922
678	QY	AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp	697
921	Db	GACGAGGTGGAGAAGCCCATCCGATGTATTCAACATTTCTCTCCAGGTGTCGAGGAC	862
698	QY	GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMet	717
861	Db	GGACGCTGACCGACAGTCACGGCGGTACGGTGGACTTCCGCAACACCGGTGGTGGATG	802
718	QY	ThrSerAsnLeuGlyAlaGluHisLeu-----LeuAlaGlyLeuThrGlyLysValThr	735
801	Db	AGCTTCAACCTCGTTCGGCGCAGATCCAGGAGCTGGCGGC-----GAC	757
736	QY	MetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeu	755
756	Db	CCGAGGCGCAACGTGCCGAGTGAAGACGGGTCAATGGCGCACTTCGTCGGAAATTC	697
756	QY	LeuAsnArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLys	775
696	Db	ATCAACCGATCGACGAAGTGTGTGTTCGAGCGCTGGTCGGAGCAGATGCCCGC	637
776	QY	ValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeu	795
636	Db	ATCGCGCGAGATCCAGCTCGCTCGCTCGCGCAAGCGCTGGCGCGCGAGCTGAGCCTG	577
796	QY	AlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGly	815
576	Db	GNACTGACGAGGAGCGCTGACAAAGCTGATTCGCTGGCTTCGACCCCGTCTATGGC	517
816	QY	AlaArgProIleArgArgTrpMetGluLysLysValValThrGluLeuSerLysMetVal	835
516	Db	GCACGCCGCTGAAGCGGCGCCATCCAGCGCTGGATCGAGAACCCCGTGGCGCACTGATC	457
836	QY	ValArgGluGluIleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeu	855
456	Db	CTGGCCGGCAATTCCGCCCGGTGCCAGTATTCGCGCAAGGTGGAGGCGGACGAGATC	397
856	QY	ValTyrArgValGluSerGlyGlyLeu-----ValAspAlaSerThrGlyLysLys	872
396	Db	GTCTT-CGCTTGACCTCTTCGGGGCGGACAGGGAAGCCCGCTTCGGCGGG-----	344
873	QY	SerAspValLeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAla	899
343	Db	-----CTTTTTCATGG-CCGTCCGGTAGGCGGAGCGTGGCGCA	303

RESULT II
US-09-199-637A-1/c

US 03-153-037A-17C
; Sequence 1, Application US/09199637A

; Patent No. 6355411

; GENERAL INFORMATION:

APPLICANT: Ausubel, Frederick M
APPLICANT: Goodman Howard M

APPLICANT: GOODMAN, HOWARD M.
APPLICANT: RAHME, LAURENCE G.

APPLICANT: Mahajan-Miklos, Shalina

APPLICANT: Tan, Man-Wah

; APPLICANT: Cao, Hui
 : APPLICANT: Drenthard Elana

APPLICANT: Drenkard, Eliana
APPLICANT: Tsongalis, John

TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
 TITLE OF INVENTION: SEQUENCES AND USES THEREOF
 FILE REFERENCE: 00786/361002
 CURRENT APPLICATION NUMBER: US/09/199,637A
 CURRENT FILING DATE: 1998-11-25
 PRIOR APPLICATION NUMBER: 60/066,517
 PRIOR FILING DATE: 1997-11-25
 NUMBER OF SEQ ID NOS: 437
 SOFTWARE: FastSeq for Windows Version 4.0
 Q ID NO 1
 LENGTH: 42235
 TYPE: DNA

ORGANISM: Pseudomonas aeruginosa
 FEATURE:
 NAME/KEY: variation
 LOCATION: (1)...(42235)
 OTHER INFORMATION: N is any nucleic acid.
 09-199-637A-1

Alignment Scores:
 i. No.: 1.37e-205 Length: 42235
 Score: 2133.00 Matches: 440
 Percent Similarity: 68.90% Conservative: 178
 Local Similarity: 49.05% Mismatches: 257
 Indels: 24
 Gaps: 8

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 26176 ATCGAATAGACCGTTGACGACGAGCTCAACTCGCGCTCTCCGACGCCAGTCCCTG 26117

21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 26116 GCCGTGGCCATGACCATCCGCCATCGGCGGTCGACCTGTTCCGCTTCCGCTGCTCGAG 26057

41 AspProThrGlyLeuPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaGln 60
 26056 CAGCAAGCGGTTGATCAAGCCCTGCTGATCAGCTCGGCTTCGATATCGCGCCCTG 25997

61 SerAlaGluArgValIleAsnGlnAlaLeuLysLeuProSerGlnSerProPro 80
 25996 CGCAGCGGC-----CTCAACAAGAACTCGACGCGCTCCGAGATCCAGAGCCGAC 25943

81 AspAspIleProAlaSerSerLeuIleLysValIleArgArgAlaGlnAlaGln 100
 25942 GCGCAGTGAACCTGTCCTCCAGGATCTCGACGCTGCTCAACGAGCTGACCGCTGGCC 25883

101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuMetGlyLeuLeu---Glu 119
 25882 CAGCAGAGGGCGACCAAGTTCATCTCCAGCAGCTGGTATGTCGCGCGATGACGAG 25823

120 AspSerGlnIleArgAspLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
 25822 AACACAGGCTCGCAAGCTGCTGTCGCGCAGGGGTGTCGCGAAGCGCTGAGAA 25763

140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159
 25762 GCGGTGGCCAACTGCTGGC-----GCGAAGCGGTGAACGACCCGACCTCGAGAG 25709

160 AsnPheGlnAlaLeuLysThrThrGlyArgAspLeuValGluGlnAla-----GlyLys 177
 25708 TCGCGCAGCGCTGGAACAAGTATACCTGTCGATCAGCATGACCAAGCGCGCGAGGCAAG 25649

178 LeuAspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArg 197
 25648 CTCGACCCCGGTGATCGTTCGACGACGAGATCCGCGGACCATCCAGCTCCGAGCGG 25589

198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217
 25588 CGGACCAAGAACAAACCCCGTGTGATCGCGGACCCGCGGTGCGGACCGGCTCGGACCCCATCGTC 25529

QY 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 DB 25528 GAGGGCTGGCCAGCGCATCATCAACGGCGAAGTGCAGGCGCTCAAGGACAGCGC 25469
 QY 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysThrArgGlyGluPheGlu 257
 DB 25468 CTGCTGGCCCTGGACATGGGGGCGCTGATCGCGGTGCCAAGTTCGCGCGAGTTCGAG 25409
 QY 258 GluArgLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPhe 277
 DB 25408 GAACGCTGAGGCGGTCTCTCAACGAACCTGGGCAAGAGGCGGCTCATCTCTGTT 25349
 QY 278 IleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAla 297
 DB 25348 ATCGAGAACTGCACCATGTGTCGCGCGCGCAAGGCGGAGGTGCCATGACCGCCG 25289
 QY 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeu 317
 DB 25288 AACATGCTCAAGCGCGCTCTGGCGCGCGCGAGCTGCACTGGTGGTGTCTACTACCTC 25229
 QY 318 GluGlyLysArgLysValGluLysAspAlaAlaPheGluArgArgPheGlnVal 337
 DB 25228 GACGAGTATCGCCAGTATCATCGAGAAAGATGCGCGCTGAGAGCGCGCTTCCAGAGGTG 25169
 QY 338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLysGluLys 357
 DB 25168 CTGGTGGAGCAACCGCGAGGAGACCATTCGCTCTCTCGTGGCTCAAGGACGC 25109
 QY 358 TyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377
 DB 25108 TATGAAGTGCACACCGGTGAGCATCACCGACGCGCGATCATCGCGCGCCCAAGCTG 25049
 QY 378 SerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 DB 25048 TGCACCGCTCATCATCACCAGTTCGCAACTGCGGACACTGCGGACAGGCTCATCGACG 24989
 QY 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGlu 417
 DB 24988 GCGCCAGCCCATCGCATCGGATCGACTTCAAGCGCGAGGAACTGGATCGTCTCGAC 24929
 QY 418 ArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAla 437
 DB 24928 CGTGGCTGATCGAGTGAAGTGCAGCGGAGCGCTGAAGAGGAGGACGACGAGGAC 24869
 QY 438 SerLysAlaArgLeuIleGluValArgLysGluLeuAspLeuArgAspLysLeuGln 457
 DB 24868 ACCAGGAGCCCTGGCGAAGCTGAGGAGGATATCGTCAAGCTCGAGCGGATACGCC 24809
 QY 458 ProLeuThrMetLysTyrArgLysGluLysGluIleAspGluIleArgArgLeuLys 477
 DB 24808 GACCTCGAGGAGATCTGAGATCGAGAGCGCGAGGTGAGGCTCGGCGCAGATCCAG 24749
 QY 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGlnAlaGluArgArgTyrAspLeuAla 497
 DB 24748 CAGAAGATCGAGGCGCAAGCAGAGATGAGGCGCGCGCGCGCAAGGCGACCTCGAG 24689
 QY 498 ArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 517
 DB 24688 AGCATGCGCGCTGAGTACCATGACCATCCCGAGCCCTGGACGACGAGTGTGTC 24629
 QY 518 GluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
 DB 24628 GACCAGCAGCGCAAGACCGAGACCATGCTGCTGCGCAACAGGTGACCGACGAGGAATC 24569
 QY 538 AlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
 DB 24568 GCGAGTGTGTTCCAAAGTGCAGCGGTATCCCGGTGTCGAAGATGCTCGAGGCGGAGCG 24509
 QY 558 GluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
 DB 24508 GAGAAGTGTGTCGCGATGGAGGAGGAGCTGCATCGCGGAGTGTGATCGCGGACGAGGCG 24449
 QY 578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuArgAlaGlnGln 597

QY	618	AlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGlu	637
Db	1864	GCCTGCGCCAGTTCCTCTCGATACCGAGGAGCGCTGCTGCGGATCGATATGTCGAG	1923
QY	638	TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHis	657
Db	1924	TTTCATGAGAAACACTCGGTGGCCGCCCTGATCGGGCGCTCTCGGCTACGTCGGCTTT	1983
QY	658	GluGluGlyGlyGlnLeuThrGluAlaValArgArgArgProTyrCysValIleLeuPhe	677
Db	1984	GAGGAAGCGGCTACTTGACCGAGGGGATCGCCGAGCCCTTACTCGGTGGTGTGCTG	2043
QY	678	AspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp	697
Db	2044	GACGAGGTGAGAAAGCCCATCCGGATGTATTCAACATTCTCTCCAGGTGCTCGAGGAC	2103
QY	698	GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMet	717
Db	2104	GACGCGCTCAGCGACAGTCACCGGGCGTACGGTGGACTTCGCAACACCGTGGTGGTATG	2163
QY	718	ThrSerAsnLeuGlyAlaGluHisLeu-----LeuAlaGlyLeuThrGlyLysValThr	735
Db	2164	ACCTCCAACTCGGTTCCGCACAGATCCAGGAGCTGGCCGGC-----GAC	2208
QY	736	MetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeu	755
Db	2209	CGCAGGGCGCAACGTCGCCGAGTGATGACGCGGTCAATGCACACTTCGTCGGAAATC	2268
QY	756	LeuAsnArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLys	775
Db	2269	ATCAACCGGATCGAGAAATGGTGATTCAGACCGCTGGCTCGCGACGATCGCCGCG	2328
QY	776	ValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeu	795
Db	2329	ATCGCGGAGATCCAGCTCGTGGCTGGCAAGCGCCTGGCGAGCGCGAGCTGAGCCTG	2388
QY	796	AlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGly	815
Db	2389	GAACTGACCGAGGAGCGCTGCAAGCTGATCGCCGTCCGCTTCGACCCCGTCTATGGC	2448
QY	816	AlaArgProIleArgArgTrrMetGluLysLysValValThrGluLeuSerLysMetVal	835
Db	2449	GCACGCCGCTGAAGCGGGCCATCCAGCGCTGGATCGAGAACCCGCTGGCGCACTGATC	2508
QY	836	ValArgGluGluIleAspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeu	855
Db	2509	CTGGCCGCCAAATTCGGCCGGTGCCTGATATCTCGCGGAAGGTGGAGGACGAGATC	2568
QY	856	ValTyr	857
Db	2569	GTCTTC	2574

```

RESULT 13
US-09-252-991A-4897/c
; Sequence 4897, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4897
; LENGTH: 2970
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4897

```

Alignment Scores:

d. No.: 2,41e-207 Length: 2970
 2132.00 Matches: 427
 69.95% Conservativeness: 176
 49.54% Mismatches: 245
 46.86% Indels: 14
 4 Gaps: 6

09-812-350-17 (1-911) x US-09-252-991A-4897 (1-2970)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 2949 ATGCGAATAGACCGTTTACACGAGCAAGCTGCAACTGGCGCTCTCCGACGCGCACTGCCCTG 2890
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 2889 GCGGTGGCCATGACCATCCGCCCATCCAGCGGCTGACCTGCTTCCGCGCTGCTCGAG 2830
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
 2829 CAGCAAGCGCGTTCGATCAAGCCCTGCTGATGAGTTCGGCTTCGATATCCGCGCTG 2770
 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
 2769 GCGAGCGGC-----CTCAACAAGAACTCGACGCGCTGCCGAAGATCCAGAGCCCGACC 2716
 81 AspAspIleProAlaSerSerSerLeuIleLysValIleArgAlaGlnAlaAlaGln 100
 2715 GCGAGGTGAACCTCTCCAGGATCTCCAGCGCTGCTCAACGAGCTGACCGCTGCGCC 2656
 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu---Glu 119
 2655 CAGCAGAAGCGGACCACTTCTCCAGCGAGTGTGTTATGCTGGCGCGATGAGACGAG 2596
 120 AspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSer 139
 2595 AACACGAGCTCGGCAAGCTGCTCGCGCAGCGCGTGTCCGCAAGCGCGCTCGAAGAT 2536
 140 GluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThr 159
 2535 GCGGTGGCCAACTCGGTGGC-----GGCGAAGCGGTGAACGACCGCGAGCTGAGGAG 2482
 160 AsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla-----GlyLys 177
 2481 TCGCGCCAGGCGCTGGCAAGTACACCGCTGCATGACCATGACCAAGCGCGCGAGAGGCAAG 2422
 178 LeuAspProValIleGlyArgAspGluLysLeuArgValValArgIleLeuSerArg 197
 2421 CTCGATCCGGTGATCGGTGCGGACGACAAATCCCGGACCATCCAGGTCCTGACGCGG 2362
 198 ArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValVal 217
 2361 CGGACCAAGAACACCCCGGTCTGATCGCGAACCCTCGCGCTCGGCAAGACCGCATCGTC 2302
 218 GluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 237
 2301 GAGGCGCTGGCCCGACGCGCATCAACGCGGAACTGCGGACCGCTCAAGGACAAAGCGC 2242
 238 LeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGlu 257
 2241 CTGCTGGCCCTGGACATGGGGCGCTGATCCCGGTCGCAAGTTCCCGCGGAGTTCGAG 2182
 258 GluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPhe 277
 2181 GAACGCTTGAAGCGAGTCTCTCAACGAACTGGGCAAGCAGGAGCGCGGTCTCTCTTC 2122
 278 IleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAla 297
 2121 ATCCAGCAACTGCACACCTGGTTCGGTCCCGCAAGCGGGAAGGTGCATGAGCGCGGC 2062
 298 AsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrLeu 317

Db 2061 AACATGCTCAAGCGCGCTCTGGCGCGCGAGCTGCACTGCGTGGTCTACTACCTC 2002
 QY 318 GluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnVal 337
 Db 2001 GACGAGTATCGCAGTACATCGAGAGGATGCGGCTGAGCGCGCTTCAGAGGTTG 1942
 QY 338 TyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLysGluLys 357
 Db 1941 CTGGTGGACGAAACCGAGCGAGGAGACACGATCGCATCTCCGTGGTCTCAAGGAACGC 1882
 QY 358 TyrGluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeu 377
 Db 1881 TATGAAGTGCACACCGGCTGAGCATCACCGAGCGGCGGATCATCGCGCGCCCAAGTGT 1822
 QY 378 SerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 397
 Db 1821 TCGACCGCTACATCACCGATCGGCAACTGCGGACAGGCCATCGACTGATCGACGAG 1762
 QY 398 AlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGlu 417
 Db 1761 GCGCGCAGCGCGATCGCATGAGATCGACTCCAAGCCAGAGAACTGGATCGTCTCGAC 1702
 QY 418 ArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAla 437
 Db 1701 CGTCCGCTGATCCAGTGAAGATCGAGCGCGAGCGCTGGAAGAGGAGACGACGAGGCC 1642
 QY 438 SerLysAlaArgLeuIleGluValArgLysGluLeuAspLeuArgAspLysLeuGln 457
 Db 1641 ACCAGGAAGCGCTGGCCAAAGCTGGAGGAGGATATCGTCAAGCTCGAGCGCAATACGCC 1582
 QY 458 ProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgGluLys 477
 Db 1581 GACCTCGAGGATATCTGGAAGTCCGAGAGGCGGAGGTGTCGAGCGCTCGCGCAGATCCAG 1522
 QY 478 GlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAla 497
 Db 1521 CAGAGATCGAGCGGCAAGCAGGAGATGAGGCGGCGCGCGGCGGCGGCGGCGGCGGCGG 1462
 QY 498 ArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 517
 Db 1461 AGCATGGCGCGCATCCAGTACCAGACCATCCCGGACCTGGAAACGCGAGCTGCAGATGTC 1402
 QY 518 GluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
 Db 1401 GACCAACGCGCAAGACCGAGAACCATGTTGCTGCGCAACAGGTGACCGACGAGGAATC 1342
 QY 538 AlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
 Db 1341 GCGAAGTGGTTTCAAGTGGACCGGTATCCCGGTGCAAGATGCTCGAGGCGGCGGCGG 1282
 QY 558 GluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
 Db 1281 GAGAGCTGCTGGCATGGAGCAGGAGCTGATCGCGAGTGTATCGGCGAGGAGCGG 1222
 QY 578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGln 597
 Db 1221 GTAGTGGCGGTGTCAACGCGCTGCGCGCTGCGCGCGCGCTCGCGATCCGCAACCGG 1162
 QY 598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLys 617
 Db 1161 CCGAGCGGTCTGTTCTTCTTCTTCTCGCGCGCGAGCGGCTGGGCGAGCGAGTTGTGCAAG 1102
 QY 618 AlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGlu 637
 Db 1101 GCGCTGGCGAGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1042
 QY 638 TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHis 657
 Db 1041 TTCATGGAGAAACACTCGGTGGCGCGCTGATCGCGCGCGCTCGCGGCTTCTCGGCTTT 982
 QY 658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe 677
 Db 981 GAGGAGGCGCGCTACTGACCGAGGCGATCCGCCCAAGCCCTACTCGGTGGTGGTGGTGG 922

678 AspGluValGluValAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAsp 697
 921 GACGAGGTGGAAGAGCCCATCCGATGATTCATCAATCTCTCCAGGTGTCGAGGAC 862
 698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleMet 717
 861 GCACGCGCTGACCGACAGTCACGCGGCTAGCGGTGGACTTCGCGCAACACACCGTGTGTGATG 802
 718 ThrSerAsnLeuGlyAlaGluHisLeu-----LeuAlaGlyLeuThrGlyLysValThr 735
 801 ACCTCCACCTCGGTTCCGACACAGATCCAGGAGCTGGCGGC-----GAC 757
 736 MetGluValAlaArgAspCysValMetArgGluValArgGlyHisPheArgProGluLeu 755
 756 CGCAGGCGCAACGTCGCGCATGATGACCGCGTCATGACATTCCTCCGCGCAATTC 697
 756 LeuAsnArgLeuAspGluLeuValValPheAspProLeuSerHisAspGlnLeuArgLys 775
 696 ATCAACCGGATCGACGAAGTGTGTATTTCGAGCGCTGGCTCGCGAGCAGATCGCCGCG 637
 776 ValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeu 795
 636 ATCGCGGAGATCCAGCTCGGTGCGCTGCGCAAGCGCTGGCGGAGCGCGAGCTGAGCCTG 577
 796 AlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGly 815
 576 GAACGAGCCAGGAGCGCTGACAAAGCTGATCGCGCTCGCGCTCGACCCCGTCTATGGC 517
 816 AlaArgProIleArgArgTyrMetGluLysValValThrGluLeuSerLysMetVal 835
 516 GCACCGCGCTGAGCGGCGCCATCCAGCGCTGGATCGAGAACCCGCTGGCGCAACTGATC 457
 836 ValArgGluGluLeuAspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeu 855
 456 GTGGCGGCAATTCGCGCGGCTGCCAGTATCTCGCGCAAGGTGGAAGGAGAGCAGATC 397
 856 ValTyr 857
 396 GTCTTC 391

UNT 14
 09-134-001C-1142
 sequence 1142, Application US/09134001C
 Patent No. 6380370
 GENERAL INFORMATION:
 APPLICANT: Lynh Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 1142
 LENGTH: 2616
 TYPE: DNA
 ORGANISM: Staphylococcus epidermidis
 09-134-001C-1142
 Alignment Scores:

d. No.:	5,54e-205	Length:	2616
re:	2108.00	Matches:	413
cent Similarity:	75.13%	Conservative:	170
t Local Similarity:	53.22%	Mismatches:	173
ry Match:	46.33%	Indels:	20
	4	Gaps:	8

 09-812-350-17 (1-911) x US-09-134-001C-1142 (1-2616)

QY 86 SerSerSerLeuIleLysValIleArgArgAlaGlnAlaGlnLysSerArgGlyAsp 105
 Db 262 AGTGCTAAACAAATGAATATTATAGCAAAAGCCGAGAAATATATGAAGTCTATGAAGAT 321
 QY 106 ThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAspSerGlnIleArgAsp 125
 Db 322 GAGTTATTCTATGGAGCATATATTACGTGCGACCATAGATAGTACTGATGAACTACTCAA 381
 QY 126 LeuLeuAsnGluValGlyValAlaLalaThrAlaArgValLysSerGluValGluLysLeuArg 145
 Db 382 -----AAATGGGTGCGCAATAAAAGTAGAAGTGTATAAGAAATTTATACTAAAGTTAGA 435
 QY 146 GlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsnPheGlnAlaLeuLys 165
 Db 436 GGG-----GGAATCATGTGACATCTCAAAATCCAGAAAGTAACTATGAAGCTTTAGAA 489
 QY 166 ThrTyrGlyArgAspLeuValGluGlnAla-----GlyLysLeuAspProValIleGly 183
 Db 490 AAATATGCTGTCATCTTGTAGAGAGTACGTCAAGGTAAATGAGCCCTGTCACTCGGT 549
 QY 184 ArgAspGluGluLeuArgValValArgIleLeuSerArgArgThrLysAsnAsnPro 203
 Db 550 AGAGATGAAGAAATAGCTAATCAATTCGTATATTAACTCGAAACAAAAATAATATCCG 609
 QY 204 ValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGlyLeuAlaGlnArg 223
 Db 610 GTTTTATCGGTGAAACGAGTGTAGTAAACTGCAATTTGTTGAAGTTTATGACCAACGA 669
 QY 224 IleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSerLeuAspMet 243
 Db 670 ATTGTACGTAAGACCGTTCAGATCATTTACTCGATAAACGATTTTCGAGCTTGATTTA 729
 QY 244 GlyAlaLeuValAlaGlyLysTyrArgGlyGluPheGluGluArgLysSerVal 263
 Db 730 AGTCCCTTAGTTTCAGGTGCAAAATATAGAGGTGAATTTTGAAGACGTTTAAAGCAGTA 789
 QY 264 LeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGluLeu 283
 Db 790 TTGAAGAAGTAAAGAACTGAAAGTCAATCTTATTATTGATGAAATTCATATG 849
 QY 284 ValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPheLysProMet 303
 Db 850 TTAGTAGGCGCTGTTAAACTGATGCTGTATGGACGAGGAAACATGTTGAACCAATG 909
 QY 304 LeuAlaArgGlnLeuArgCysIleGlyAlaThrThrLeuGluGluTyrArgLysTyr 323
 Db 910 CTTGCGCGTGTGAATTCATTTGATTTAGTGTCTAGCACATTAATCAATATCTGTAATAT 969
 QY 324 ValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyrValAlaGluProSer 343
 Db 970 ATGAAAAGATTCAGCATTAGAAGTCTGTTTCCAAAAGTTGGGGTCAAGTGAACCTGAT 1029
 QY 344 ValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGluGlyHisGly 363
 Db 1030 GTTGAGATACAAATATCAATTTTACGTGCTCTAAAGAACGTTATGAGTCTACCATGTT 1089
 QY 364 ValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArgTyrIleThr 383
 Db 1090 GTACGTATACAAAGCCGTGCACTTGTTCAGCGGCTGAGTATCTGTATGATATATATACC 1149
 QY 384 GlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAlaAsnValArg 403
 Db 1150 GATAGATTTTCCAGATTAAGCTATAGATTTAGTCGACCAAGCTTTGGCCACATACGT 1209
 QY 404 ValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArgMetGlnLeu 423
 Db 1210 ACAGAAATGGGATCAATCCACAGAGTTAGACCAAGTAAATCGTCGAGTAAATGCAATTA 1269
 QY 424 GluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAlaArgLeu 443
 Db 1270 GAAATTCGAAGAAAGCGCTTCGAAATGAATCTGATAACGCAAGTAAACATCGTTTGA 1329
 QY 444 GluValArgLysGluLeuAspLeuArgAspLysLeuGlnProLeuThrMetLysTyr 463

1330 GAGCTTCAAGAGAAATTATCTAATGAAAAAACAATCTCTTTTAAATCACTGTT 1389
464 ArgLysGluLysGluArgLysGluArgLysGluArgLysGluArgLysGlu 483
1390 GAACAAGAAAAAAGAAAAA-----ATTGCTAAAGTACAAAGAAAAAGCTGCTGAAT 1440
484 MetPheSer-----LeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAla 500
1441 GATAGTAGTCTGCTCAAGCTCTTGAAGATGCACAAACAGAGAGTAATCTTCAAAAAGCAGCA 1500
501 AspLeuArgTyrGlyAlaLeuGlnGluValGluSerAlaLeuAlaGlnLeuGlu 518
1501 GAATTACAATATGTTCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCA 1560
519 -----GlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGlu 535
1561 TTTCAGATGAATCGCGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1620
536 HisLeuAlaGluValValSerArgTyrThrGlyLeuProValThrArgLeuGlyGlnAsn 555
1621 GAAATAGGCGACATAGTATGATCAATGCAGCAGTATACCTGTATCAAACTTGTGGAACA 1680
556 GluLysGluArgLeuLeuGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsn 575
1681 GAAAGAGAAAGCTATTAAAGTTTAAAGTGAATCTTACACAAACGCTAGTAGTCAAGAT 1740
576 GlnAlaValAsnAlaValSerGluAlaLeuLeuSerArgAlaGlyLeuGlyArgAla 595
1741 AAAGCAGTTGATTAGTATGATCAGCAGCAGTATGAGCAGCAGTATGAGCAGCAG 1800
596 GlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeu 615
1801 AATAGACCAATCGAAGTTCTTATTTCTTAGCAGCTTACTGGAGTAGGTAATACTGAATTA 1860
616 AlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgLysLeuMet 635
1861 GCAAAATCGCTTCATCATCTTTTCGATCTCGAATTCGAAACATATGATTAGATTGATG 1920
636 SerGluTyrMetGluGlnHisSerValSerArgLeuLeuGlyValAlaProProGlyTyrVal 655
1921 AGCGATATATGAAAAACATGCTGTATCAGCTTTAATTTGGTGCACCTCCAGCTTATGTA 1980
656 GlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValle 675
1981 GGTCCAGATGAAGTGGTCAATTAACCTGAAGCAGTTAGACGTAATCCATCTCAGTTAT 2040
676 LeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuLeuValLeu 695
2041 TTGTTAGACGAAGTTGAAAAAGCAGCAGTATGTTTAAATGTTATTTACTTCAAACTACTA 2100
696 AspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValle 715
2101 GATGAAGTCTGCTTACGATCTTAAGAGTAGAGTGTGACTTTAAATAACCATATATC 2160
716 IleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeu-----ThrGlyLys 733
2161 ATCATGACTAGTAATATGTTTCAACAGTATTACTTGAATAATGTAAGATGCTGCTGTA 2220
734 ValThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPheArgPro 753
2221 ATTAGTGATGATACAGAGAAA---GCAGTTATGGACAGTCTACATGCTACTTCAAACT 2277
754 GluLeuLeuAsnArgLeuAspGluLeuValValPheAspProLeuSerHisAspGlnLeu 773
2278 GAAATATTAAATCGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2337
774 ArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArgLysVal 793
2338 AGTATGATTTAGTAAATTTTAAACAAATTTAAATATGAGATTTATGATCAACATATC 2397
794 AlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProVal 813

Db 2398 TCAATTGAAGTGACAGAGAGCGGAAAAAATGGCTAGGTGAAGAGCGTATGACACCAA 2457
Qy 814 TyrGlyAlaArgProfileArgArgTyrMetGluLysLysValValThrGluLeuSerLys 833
Db 2458 TTTGCTGACAGACCATTAACACGCTTTGTTTCAACGACAAATAGAACTCCAATTGCACGT 2517
Qy 834 MetValValArgGluGluIleAspGluAsnSerThrValTyrIleAsp 849
Db 2518 ATGATGATTAAAGAAAGTCTACCTGAAGGTACATAATAATTAAAGTAGAT 2565
RESULT 15
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffrs, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09198, 452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (45001)..(60000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (60001)..(75000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (75001)..(90000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (90001)..(105000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (105001)..(120000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (120001)..(135000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (135001)..(150000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (150001)..(165000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (165001)..(180000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (180001)..(195000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (195001)..(210000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
; LOCATION: (210001)..(225000)
; OTHER INFORMATION: n=a or c or g or t

NAME/KEY: misc feature
 LOCATION: (225001)..(240000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (240001)..(250000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (255001)..(270000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (270001)..(285000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (285001)..(300000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (300001)..(315000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (315001)..(330000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (330001)..(345000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (345001)..(360000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (360001)..(375000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (375001)..(390000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (390001)..(405000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (405001)..(420000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (420001)..(435000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (435001)..(450000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (450001)..(465000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (465001)..(480000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (480001)..(495000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (495001)..(510000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (510001)..(525000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (525001)..(540000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (540001)..(555000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (555001)..(570000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (570001)..(585000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature

LOCATION: (585001)..(600000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (600001)..(615000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (615001)..(630000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (630001)..(645000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (645001)..(660000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (660001)..(675000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (675001)..(690000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (690001)..(705000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (705001)..(720000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (720001)..(735000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (735001)..(750000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (750001)..(765000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (765001)..(780000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (780001)..(795000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (795001)..(810000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (810001)..(825000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (825001)..(840000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (840001)..(855000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (855001)..(870000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (870001)..(885000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (885001)..(900000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature
 LOCATION: (900001)..(915000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc feature

Alignment Scores:

Pred. No.:	1,11e-180	Length:	1230025
Score:	1912.00	Matches:	404
Percent Similarity:	66.24%	Conservative:	171
Best Local Similarity:	46.54%	Mismatches:	271
Query Match:	42.02%	Indels:	22

```
09-812-350-17 (1-911) x US-09-198-452A-1 (1-1230025)
4 Gaps: 8
4
188339 GAGAAATTTTCGATGCTCTCTCTGAGCTTTAGAGAGGCTTTTCGAACTTCTCTAAATCT 188398
24 AlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLeuLeuSerAspProThr 43
188399 TCGAACATACCTATGTCACAGAAATACCTATTACTGGCTTTATTAGAAATACAGAG 188458
44 GlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAlaAlaGlnSerAlaGlu 63
188459 TCTCTTTTATTGTTAAATTAAGACATTCATGGGAACCTCGGT----- 188503
64 ArgValIleAsnGlnAlaLeuLeuValLeuProSerGlnSerProPro----- 79
188504 ---TGCTCAATACGGCAGTTAAAGTGGCTCTCACGAGAGCCGACTGTAGTTGAAGGA 188560
80 ProAspAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAlaAla 99
188561 GAGGTGGATCCTAAACCTCTCCGGGTTTACAAACCTCTTTAGGGATGCCAAACAGAG 188620
100 GlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGlu 119
188621 GCAAAGACATTAGGAGATGAATACATTTCTGGAGATCATCTGCTGCTTTTGGAGT 188680
120 AspSerGln-----IleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgVal 137
188681 TCAACAAAGAGCCCTTTAATTTCTTGGAGCAACACAAAGTTAGTTTAAAGATCTT 188740
138 LysSerGluValGluLysLeuArgGlyLysGluLysLysValGluSerAlaSerGly 157
188741 AAGAATCTGATTACTAAATACGA-----CGAGGAATCGTATGATTCGCCAAGCGCT 188794
158 AspThrAsnPheGlnAlaLeuLysThrTyArgAspLeuValGluGlnAla----- 175
188795 GAAAGTAATTTTCAGGGTTTGAAGAAGTATTTCAAAATTTTAAACAGCATTAGCTCGTGA 188854
176 GlyLysLeuAspProValIleGlyArgAspGluGluIleArgValValArgIleLeu 195
188855 GGTAACTGGATCCTGTGATCGTAGAGATGAAGAAATTCGTAGAACCATCCAGTGTCT 188914
196 SerArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAla 215
188915 TCCCGTAGAACTAAATAAACCTATGCTATTGCTGAGCCGGGTGAGGAAACTGCT 188974
216 ValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAsp 235
188975 ATAGCAGAGAGGATAGCTCTTAGGCTTATCCAGGGTGTATTCCTGAAATCTCTCAAAGGT 189034
236 ValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyValAlaLysTyArgGlyGlu 255
189035 AAACAGCTTTATGCTTAGATATGGAGCTTTGATTGCGAGGAGCTAAGTATCGAGGTGAG 189094
256 PheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIle 275
189095 TTTGAGAAAGACTAAAGAGTGTTTTAAAGATGTAGATCTCGAGATGGCGAGCACATT 189154
276 LeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAsp 295
189155 ATCTTTATTGATGAGTGCTACTCTTTGTTGAGCAGGAGCTACTGTATGGAGCTATGAT 189214
296 AlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThr 315
189215 GCTGCAATCTTTAAAGCTTCATTAGCAAGAGGAGCGCTACACTGTATTGGCGGAGC 189274
316 ThrLeuGluGluTyArgLysTyArgValGluLysAspAlaAlaPheGluArgPheGln 335
189275 ACTTTGAATGAGTATCAGAGATATATGAAAGATGCTGCTTTTGAACGTCGATTTTTCAG 189334
336 GlnValTyValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLys 355
```

```

189335 CCTATTTTGTGACAGAGCCTCTTTTGGAGATGCTCTTTTATTCTTCTGTGACTAAGA 189394
356 GluLysTyArgGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuLeuAsnAlaAla 375
189395 GAAATAATATGAAATTTTCCATGGAGTCAGGATTCAGAGGGGCTTTGAAATGCCGCGATC 189454
376 GlnLeuSerAlaArgTyIleThrGlyArgHisLeuProAspLysAlaIleAspLeuVal 395
189455 CTACTTTCTCTATCGTTATATCCAGATCGCTTTCTTCCAGTAAGCTATCGATTGTGATA 189514
396 AspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsn 415
189515 GATCAAGCGGCAAGTTTAATTCGATGCAAAATGGTAGTCTTCTCTCTCTTATGATGAA 189574
416 LeuGluArgLysArgMetGlnLeuIleGluLeuHisAlaLeuGluArgGluLysAsp 435
189575 AAGGAGAGAGAGCTGCTGCTTTGATCGTTTANGCAAGAGCTATAAACCGAGCAATCT 189634
436 LysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLys 455
189635 CTTTCTTATCAAGAAGAGCGGATGCTATGCAGAGTCTATAGATGCTTTTGACAGAGGAA 189694
456 LeuGlnProLeuThrMetLysTyArgLysGluLysGluArgIleAspGluIleArgArg 475
189695 TTAGCATCTCTACGTTTGGGTTGGGATGAAGAAGAAGTTGATTTCCGGGCTCAAGGAA 189754
476 LeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyAsp 495
189755 AAAAGAAATTCCTTGGAAAGTATGAATTTTCTGAAGAGAGCGGAGCGTGTTCAGAC 189814
496 LeuAlaArgAlaAlaAspLeuArgTyGlyValAlaIleGlnGluValGluSerAlaIleAla 515
189815 TATAATCGTGTAGCTGAGCTTCGTTATAGTTTAAATTTCCCAACTTGAAGAAGAAATCAAA 189874
516 GlnLeuGluGlyThr---SerSerGluGluAsnValMetLeuThrGluAsnValGlyPro 534
189875 CAGGATGAACCTCTTTAAATCAAGAGATAACCGTCTCTTCAAGAAGAAGTTGACGAG 189934
535 GluHisIleAlaGluValValSerArgTyThrGlyIleProValThrArgLeuGlyGln 554
189935 CGATTGATTCGCAAGTGTGTAGCTAATTCGACAGGGATTCCTGTGCAGAAAATGCTAGAA 189994
555 AsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGln 574
189995 GCGGAGCTGAGAAACTGTTAAATCTTGAAGAATCCTTAGAAGACGCTGTGTAGGACAG 190054
575 AsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArg 594
190055 CTTTTCAGCTCTCTCGGTTAGTGTCTTATTCGTGCTGCACGTTAGGTTTAAATGAT 190114
595 AlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGlu 614
190115 CCTCAAGCTCCTTAGAGTCTTTTATTTTAGGCGCAACAGGGGTAGAAAACCGAG 190174
615 LeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAsp 634
190175 CTTGCAAAAGCTCTTCAGATCTCTTTTCAATAAAGAGAAAGCTATGTCGCTTCGAT 190234
635 MetSerGluTyMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTy 654
190235 ATGTCAAGTATATGAAAGCAATCCATTTCCAGCTTATAGGATCTTCTCCAGGGTAT 190294
655 ValGlyHisGluGluGlyGlnLeuThrGluAlaValArgArgProTyCysVal 674
190295 GTGGTTATGAGGAAGTGGAGTCTTCTGAGGCTCTTCGACGAGCTCCCTATTTCAGTA 190354
675 IleLeuPheAspGluValGluLysAlaHisAlaValaPheAsnThrLeuLeuGlnVal 694
190355 GTTCTCTTTGATCAGATGAGAAAGCAGATAAGGAAGTTCTAAATATCTCTTTTACAGGT 190414
695 LeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgSerVal 714
```


190415 TTTGATGATGGGATTCTTACGGATGGGAAAAACGCAAGTAATTTGTAATAATGCCCTTG 190474
715 IleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysVal 734
|||||
190475 TTTATCATGACATCAATATAGTCTCCAGAA---CTTGCAGATTATTGTTCAAAAAA 190531
735 ThrMetGluValAlaArgAspCysValMetArgGluVal-----ArgLysHis 750
|||||
190532 GGAAGTGGAGCTTACGAAAGACGATTCTTCTCTAGTCTCTCCAGTATTGAAAAGATAC 190591
751 PheArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspProLeuSerHis 770
|||||
190592 TTGAGCCCTGAATTTATGACCGAATTGATGAGATCTCTCTTTTGTTCATTAAACGAAA 190651
771 AspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGlu 790
|||
190652 GAAGATATCGTGAATAATAGTTGGCATTCAATGCGAAGGATTGCCAGAGATTAAAGCCA 190711
791 ArgGlyValAlaLeuAlaValThrAspAlaLeuAspTyrIleLeuAlaGluSerTyr 810
|||
190712 CGCGGATCAATTTATCTTGGGATGATTCTGTAATATTATTTCTTAGTGAACAGGGTTAT 190771
811 AspProValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGlu 830
|||
190772 GACAGTCTTTCGGAGCCGCCCTTTAAACGTTTGATCCCAACAAAAGTTGTGATCTTG 190831
831 LeuSerLysMetValValArgGluGluIleAspGluAsnSerThrValTyrIleAspAla 850
|||
190832 CTTTCTAAGGCTTTGCTTAAAGAGATATTAAACCTGATACATCGATTGAGTTGACGATG 190891
851 GlyAlaGlyAspLeuValTyrArg 858
|||
190892 GCAAAAGAGGTGCTCGTATTATAA 190915

rch completed: February 13, 2004, 08:40:35
time : 13332 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - nucleic search, using frame_plus p2n model

on: February 13, 2004, 04:00:27 ; Search time 713 Seconds
(without alignments)
4706.577 Million cell updates/sec

le: US-09-812-350-17

fect score: 4550

uence: 1 MNPEKFKTKNETIATAHEL.....VKMKRIEIEDDDNEEMIED 911

ring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

arched: 2449703 seqs, 1841816367 residues

al number of hits satisfying chosen parameters: 4899406

imum DB seq length: 0

imum DB seq length: 2000000000

t-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

mand line parameters:

DBL=frame+p2n.model -DEV=xlp
/cgcn2_1/USPTO_spool_P/US09812350/runat_10022004_101517_13444/app_query.fasta_1.1095
Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
DCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
TRANS=human40.cdi -LIST=45 -DALIGN=200 -THR SCORE=pct -THR MAX=100
MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
KLEN=2000000000 -USAP=US09812350 @CGN 1.1.85 @runat_10022004_101517_13444
PU=6 -ICPU=3 -NO WARN -LARGESQ=0 -NEG SCORES=0 -WAIT -DSBLOCK=100
NGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
APOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

abase : Published Applications NA:
1: /cgcn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgcn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgcn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgcn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgcn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgcn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgcn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgcn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgcn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgcn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgcn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgcn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgcn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
14: /cgcn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
15: /cgcn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
16: /cgcn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
17: /cgcn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
18: /cgcn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Query	Match	Length	ID	Description
No.	Score	Match	Length	ID

1	4545	99.9	2736	10	US-09-938-842A-235	Sequence 235, App
2	4545	99.9	2736	12	US-09-938-842A-235	Sequence 235, App
3	3657	80.4	3942	12	US-10-310-154-113	Sequence 113, App
4	2315	50.9	2643	12	US-10-369-493-43734	Sequence 43734, A
5	2288	50.3	2619	14	US-10-369-493-26449	Sequence 26449, A
6	2288	50.3	2619	14	US-10-047-260-37	Sequence 37, Appl
7	2288	50.3	3180	12	US-10-310-154-116	Sequence 116, App
8	2282	50.2	2637	15	US-10-156-761-7220	Sequence 7220, Ap
9	2282	50.2	9025608	15	US-10-156-761-1	Sequence 1, Appli
10	2264.5	49.8	2571	12	US-10-369-493-44175	Sequence 44175, A
11	2247	49.4	2586	12	US-10-369-493-43449	Sequence 43449, A
12	2245.5	49.4	43980	12	US-10-398-221-5	Sequence 5, Appli
13	2245.5	49.4	3011208	12	US-10-398-221-2058	Sequence 2058, Ap
14	2234.5	49.1	2715	12	US-10-369-493-46242	Sequence 46242, A
15	2228	48.9	2588	12	US-10-369-493-44640	Sequence 44640, A
16	2225	48.9	2507	12	US-10-369-493-35669	Sequence 35669, A
17	2224	48.9	2580	12	US-10-369-493-40521	Sequence 40521, A
18	2214.5	48.7	2721	12	US-10-369-493-27677	Sequence 27677, A
19	2204	48.4	2586	9	US-09-815-243-9990	Sequence 9990, Ap
20	2199.5	48.3	2574	12	US-10-369-493-34295	Sequence 34295, A
21	2198	48.3	2577	12	US-10-369-493-35080	Sequence 35080, A
22	2196.5	48.3	2529	12	US-10-369-493-31518	Sequence 31518, A
23	2194	48.2	2574	13	US-10-301-997-46	Sequence 46, Appl
24	2194	48.2	4248	12	US-10-310-154-115	Sequence 115, App
25	2193	48.2	2574	9	US-09-815-242-6161	Sequence 6161, Ap
26	2193	48.2	2574	10	US-09-960-428-20	Sequence 20, Appl
27	2193	48.2	2574	12	US-10-369-493-47195	Sequence 47195, A
28	2184.5	48.0	2574	12	US-10-369-493-41513	Sequence 41513, A
29	2182.5	48.0	2697	12	US-10-369-493-26456	Sequence 26456, A
30	2174.5	47.8	2574	12	US-10-369-493-39286	Sequence 39286, A
31	2166.5	47.6	2547	12	US-10-369-493-39663	Sequence 39663, A
32	2166.5	47.6	2547	12	US-10-369-493-40020	Sequence 40020, A
33	2162.5	47.5	2571	9	US-09-815-242-7037	Sequence 7037, Ap
34	2162.5	47.5	1830121	12	US-10-329-670-1	Sequence 1, Appli
35	2162.5	47.5	1830121	15	US-10-329-670-1	Sequence 1, Appli
36	2158.5	47.4	2586	12	US-10-369-493-41201	Sequence 41201, A
37	2158.5	47.4	3000	12	US-10-310-154-117	Sequence 117, App
38	2151.5	47.3	2556	12	US-10-369-493-31783	Sequence 31783, A
39	2150.5	47.3	3309400	10	US-09-738-626-1	Sequence 1, Appli
40	2148.5	47.2	2571	12	US-10-369-493-32853	Sequence 32853, A
41	2146.5	47.2	2571	12	US-10-369-493-33061	Sequence 33061, A
42	2145	47.1	2556	10	US-09-738-626-3499	Sequence 3499, Ap
43	2141.5	47.1	2607	9	US-09-815-242-4175	Sequence 4175, Ap
44	2141.5	47.1	2649	9	US-09-815-242-8428	Sequence 8428, Ap
45	2141	47.1	2436	12	US-10-369-493-43090	Sequence 43090, A

ALIGNMENTS

RESULT 1

US-09-938-842A-235
; Sequence 235, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krepe, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 235
; LENGTH: 2736
; TYPE: DNA

ORGANISM: Arabidopsis thaliana
09-938-842A-235

Alignment Scores:

Seq. No.: 0 Length: 2736
Matches: 910
Conservative: 0
Mismatch: 1
Indels: 0
Gaps: 0

09-812-350-17 (1-911) x US-09-938-842A-235 (1-2736)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
1 ATGAATCCAGAGAAATTCACACACAGACAAACGACCAATTCCTACAGCTCATGAGCTA 60
21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuSer 40
61 GCTGTGAATGCAGACATGCTCAATTCCTCCTTTGCAATTTAGCTGCTTTGATCTCT 120
41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaGln 60
121 GATCCACCGGTATATTTCTCAAGCAATCTCTAGTCCGCGTGGCGAGAACGAGCTCAA 180
181 TCTGCTGAAAGAGTATCAATCAAGCCTTGAAGAGCTCTTCAATCTCTCCACCT 240
81 AspAlaIleProHisSerSerLeuIleLysValIleArgAlaGlnAlaGln 100
241 GATGATATTCACGAGTCTTAGTCTTATTAAGTCAATTCGTCGCTCAAGCTGCTCAG 300
101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuGluAsp 120
301 AGTCACGAGGTGATCTCATTTGGCTTGTGACCATTTGATTTAGTCTTCTTTGAGAT 360
121 SerGlnIleArgAspLeuAsnGlnValGlyValAlaThrAlaArgValLysSerGlu 140
361 TCTCAATCAGGATTTGTGAAGAGTCCGCTGTAGCGACGCGAGGTAAGTCTGAG 420
141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
421 GTTGAGAAGCTTCGTGGGAAAGAGGAAAGTGTGAGAGTCTTCAAGGACACAAAT 480
161 PheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
481 TTTCAGCTTTAAAGACTTATGGAAGAGATTTGGTTGACCAAGCGAGGAGCTTGATCCT 540
181 ValIleGlyArgAspGluGluIleArgArgValValArgileLeuSerArgArgThrLys 200
541 GTGATTTGGTCGTATGAGGAGATTTAGAGAGTCTGAGGATTTCTTTCAGGAGAGAACG 600
201 AsnAsnProValIleuIleGluProGlyValGlyLysThrAlaValIleGluGlyLeu 220
601 AACATCCCTGCTTATTCGAGAGCCAGAGTGTGTAACACAGCTGTGTTGAAGGTTTA 660
221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
661 GCACAAAGATTTGGAAGAGATGTGCCACACAGCTCTTACTGATGTGAGATTAATTCG 720
241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrglyArgGlyGluPheGluGluArgLeu 260
721 TTGGACATGGTGGCTTAGTGTGTGTGCTAAATACCGAGGAGAGGTTTGAAGAGGTTG 780
261 LysSerValLeuLysGluValGluAspAlaGluLysValIleLeuPheIleAspGlu 280
781 AAATCTGTTTGAAGAGAGTTCAGGACGCTGAAGGCAAGTGAATCTCTTTATTCATGAG 840
281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300
841 ATTCATTTGGTCTTGGCTGGCAAACTGAAGGTCATGAGTGCAGCTAAATCTGTTTC 900

301 LysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGluTyr 320
Db AAGCCCATTTAGCTTAGAGGCACCTTCGATGCTATGCTGTACACGCTTGAAGATAC 960
321 ArgLysTyrrValGluLysAspAlaPheGluArgArgPheGlnGlnValTyrValAla 340
Db AGGAAATATGTTGAGAAAGATGCTGCTTTGAGAGGAGGTTCCAAACAGTCTATGTTGCG 1020
341 GluProSerValProAspThrIleSerIleLeuArgGlyLysGluLysTyrrGluGly 360
Db GAGCCAAAGTGTGCTGACACCATTTAGTATCTTTAGAGGACTCAGAGGAAAGTATGAGGGA 1080
361 HisIleGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAlaArg 380
Db CATCATGCTGCGAATCCAAAGACAGAGCTCTTATAATGCTGCTCAGCTGTCTGCTCGT 1140
381 TyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCysAla 400
Db TACATACTGCTGCGCATTTACCGATAAAGCAATTCATTTGGTTGATGAGCTTTGTGCG 1200
401 AsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLysArg 420
Db AATGTGAGGTCCAGCTGTGATGATCAACCTGAAGAGATTCATTAACCTTGAAGGAAAGG 1260
421 MetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSerLysAla 440
Db ATGCAGCTGGAATTTGAATTCAGCCCTTGGAAAGGAGAGGATTAAGCCAGCAAGCT 1320
441 ArgLeuIleGluValArgLysGluLeuAspLeuArgAspLysLeuGlnProLeuThr 460
Db CGACTTATAGAGGTGCGGAAAGAGCTTGTATGATCCTGAGACACAAAGCTTCAAGCTCTCACG 1380
461 MetLysTyrrArgLysGluLysGluValArgIleAspGluIleArgGluLysGlnLysArg 480
Db ATGAATAACGAAAGGAGAGAGAGATTTGATGAGATTCGAAGGCTTAAACAGAAAGA 1440
481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrrAspLeuAlaArgAlaAla 500
Db GAAGAGCTCATGTTTTCTTTGAGGAGGACAGAGATATGACCTTGCACAGAGCTGCT 1500
501 AspLeuArgTyrrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluTyr 520
Db GATCTAAGATATGCGCAATTCAGAAAGTGAATCTGCAATTTGCCCAACTTGAAGGAACT 1560
521 SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluVal 540
1561 TCTTCTGAAGAGATGTGATGCTCAGAAAGAGGTTGGGCTGAAACATTTGCTGAGGTT 1620
541 ValSerArgTyrrThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeu 560
Db GTGAGCGTTTGGACAGGATTTCCAGTGCAGAGACTTGGCCAAATTCAGAAAGGAGAGGTTG 1680
561 IleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAlaAla 580
1681 ATTGGCTTCTGATAGTGTGATAGCGGTTGTGGGAGGAGTGTGAGGAGTAAATGCA 1740
581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
Db GTTCTGAGGCAATTTCTAAGGTCAGGCGAGGACTTGGAAAGGCCACACAGCAACTGGA 1800
601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
Db TCATTTCTTATTTCTTGGACCACTGCTGTTGGCAAACTGAGCTGCCAAGGCTCTTGTCT 1860
621 GluGlnLeuPheAspAspGluAsnLeuValArgValGlyGlnAsnGlnAlaValAlaAla 640
1861 GAGCAGCTGTTTGTATGATGAAACCTCTTAGTTCGATTCGATTCGATTCGATTCGAA 1920
641 GlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHisGluGluGly 660
Db CAACACTCTGCTCTCTGCTTCATTTGGGGGACACACAGGATGATGTTGTCACGAGGAGGT 1980
661 GlyGlnLeuThrGluAlaValArgArgProTyrrCysValIleLeuPheAspGluVal 680

```

1381 GGACAACTAACTGAGGCTGTGAGGAGCGACCTTATTGTGTCACTACTTTTGATGAGTG 2040
681 GluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspGlyArgLeu 700
2041 GAGAGGCTCAATGTGCTCTTCAACACTCTGCTCAAGTTTGGATGATGTCGATG 2100
701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSerAsn 720
2101 ACAGCGGCGAAGCAGGACAGTCAATTCAGGAACTCGGTGATTAATCATGACATCAAC 2160
721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740
2161 CTTGGTGTCAACACCTCTTGCAGGGCTAACTCGGAAAGTAACAATGGAAGTGGCCGG 2220
741 AspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAsp 760
2221 GACTGTGTGATGCGGAGGTGAGAAACACTTCAGACGAGCTCTTGAACAGGCTTGAC 2280
761 GluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGln 780
2281 GAGATTGTGTGTTTCGACCCCTTTCATGACCACTGAGGAAAGTAGCTCGGCTCAA 2340
781 MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla 800
2341 ATGAAGACGTTGCTGCTCGGCTTGTGAAAGAGGAGTGTCTTGGCAGTCACTGATGCT 2400
801 AlaLeuAspThrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArg 820
2401 GCTTTGACATATCTTTGGCAGAGATTGACCCGCTGATGCTGTCGCTTATAAGG 2460
821 ArgTrpMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIle 840
2461 AGATGGATGAGAGAAAGTGTGACAGAACTGTCAAAGATGTTGTGCGTGAAGAAATC 2520
841 AspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu 860
2521 GATGAAACTCCACTGTTTACATAGATGCGGCGCTGCTGATCTTGTGTACCGGCTAGAA 2580
861 SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla 880
2581 AGTGAGGCTCTAGTGGACGCTTCAACAGGCAAGAGTCAGATGCTGATTCATATGCT 2640
881 AsnGlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIle 900
2641 AACGGCCAAAGAAAGTGTGACGCTCAGCGGTGAAGAGATGAGATCGAGGAATA 2700
901 GluAspAspAsnGluGluMetIleGluAsp 911
2701 GAAGATCAGCATATGAGGAAATGATCGAGAT 2733

```

JUL 2

us-09-812-350-17.p2n.rnpb
 Publication No. US20040009476A9
 Application No. US/0938842A

GENERAL INFORMATION:

APPLICANT: Harper, Jeff

APPLICANT: Kreps, Joel

APPLICANT: Wang, Xun

APPLICANT: Zhu, Feng

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

FILE OF INVENTION: SAME, AND METHODS OF USE

FILE REFERENCE: SCRIPI300-3

CURRENT APPLICATION NUMBER: US/09/338,842A

CURRENT FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227,866

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264,647

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300,111

PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379

30 ID NO 235

LENGTH: 2736
 TYPE: DNA
 ORGANISM: Arabidopsis thaliana
 US-09-938-842A-235

Alignment Scores:

Pred. No.: 0 Length: 2736
 Score: 4545.00 Matches: 910
 Percent Similarity: 92.88% Conservatives: 0
 Best Local Similarity: 99.89% Mismatches: 1
 Query Match: 99.89% Indels: 0
 DB: 12 Gaps: 0

US-09-812-350-17 (1-911) x US-09-938-842A-235 (1-2736)

```

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
DB 1 ATGAATCCAGAGAAATTCACACAAAGACAAACAGAGCAATTTGCTACAGCTCATGAGCTA 60
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
DB 61 GCTGTGATGCGAGGACATCTCAATTCCTCTTGTGCAATTTAGCTGGTGTCTTGTATCTCT 120
QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGluAsnAlaAlaGln 60
DB 121 GATCCACCGGTATATTTCTTCAAGCAATCTCTAGTGGCGGTGGCGAGAACGAGCTCAA 180
QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro 80
DB 181 TCTGCTGAAGAGTGTATCAATCAAGCTTGAAGAGCTTCTTCAATCTCTCCACT 240
QY 81 AspAspIleProAlaSerSerLeuIleLysValIleArgArgAlaGlnAlaAlaGln 100
DB 241 GATGATATTCAGCAGGATCTAGTCTTATTAAGTCAATTCGCTGCTCAAGCTCTCTCAG 300
QY 101 LysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuAsp 120
DB 301 AAGTCACGAGGTGATCTCAATTTGGCTGTTCACCAATTTGATTTGGTCTTCTTCAAGAT 360
QY 121 SerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGlu 140
DB 361 TCTCAAAATCAGGAGTATTGTTGAACGAGTCTGCTGTAGCGAGCGGCGAGGCTTAAAGTCTGAG 420
QY 141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
DB 421 GTTGAAGAGCTTCGTGGAGAGAGAGAGAGAGAGTTGAGAGTCTTCAGGGGACACAAAT 480
QY 161 PheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeuAspPro 180
DB 481 TTTCAAGCTTTAAAGACTTATGGAAGAGATTGGTTGAGCAAGCAGGAGGAGCTTATCTCT 540
QY 181 ValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArgThrLys 200
DB 541 GTGATTTGGTCTGTGATGAGAGATTAGAAGAGTCTGAGGAGTTCTTTCGAGGAGAACGAAG 600
QY 201 AsnAsnProValLeuIleGlyLysProGlyValGlyLysThrAlaValValGluGlyLeu 220
DB 601 AACATCTCTGCTTATTGGAGGACGAGGTTGTGTAACAGCTGTGTTGAAGGTTTGAAGGTTT 660
QY 221 AlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIleSer 240
DB 661 GCACAAAGGATTTGTAAGAGGAGATGTGCCAACAGTCTTACTGATGTGAGATTAAATTCG 720
QY 241 LeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluArgLeu 260
DB 721 TTGGACATCGGGTCTTATTGCTGCTTAAATCCGAGGAGAGTTTGAAGAAAGGTTG 780
QY 261 LysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAspGlu 280
DB 781 AAATCTGTTTGAAGAGAGTTGAGGACCTGAGGCAAGTGTATCTTTTATTGATGAG 840
QY 281 IleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPhe 300

```

Qy	661	GlyGlnLeuThrGluAlaValaArgArgProTyrCysValIleLeuPheAspGluVal	680
Db	1981	GGACAACCTAACTCAGGCTGTGAGGAGCGACCTTATGTGTCATACTCTTTGATGAAGTG	2040
Qy	681	GluIysAlaHisValAlaValPheAsnThrLeuLeuGlnValIleuAspGlyArgLeu	700
Db	2041	GAGAAGCCTCATGTTGCTGCTTCAACACTCTGCTCCAAGTTTGGATGATGTCGATTG	2100
Qy	701	ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSerAsn	720
Db	2101	ACAGACGGCAAGCGCAGGACAGTCGATTCAGAACTCGTGATATCATGCATCAAC	2160
Qy	721	LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg	740
Db	2161	CTTGGTGCTCAACACACTCTCTGAGGGCTTAACTCGGAAAGTAACAATGGAAGTGGCCCG	2220
Qy	741	AspCysValMetArgGluValArgIysHisPheArgProGluLeuLeuAsnArgLeuAsp	760
Db	2221	GACTGTGTGATCGGGAGGTGAGAAACACTTCAGACCAGAGCTCTTGAAACAGGCTTGAC	2280
Qy	761	GluIleValValPheAspProLeuSerHisAspGlnLeuArgIysValAlaArgLeuGln	780
Db	2281	GAGATTGTGGTGTTCACCCCTTTCACATGACCAGTTGAGGAAAGTAGCTCGGCTTCAA	2340
Qy	781	MetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAla	800
Db	2341	ATGAAAGAGCGTTGCTCTCGGGCTTGCTGAAAGAGGAGTTGCTTTGGCAGTCACTGATGCT	2400
Qy	801	AlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrClyAlaArgProIleArg	820
Db	2401	GCTTTGGACTATATCTTGGCAGAGGTTATGACCCGGTGTATGTGCTAGGCGTATAAGG	2460
Qy	821	ArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGluGluIle	840
Db	2461	AGATGATGGAGNAGAGGTGGTGACAGAACTGTCAAGATGGTTGTGGGTGAGGAAATC	2520
Qy	841	AspGluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGlu	860
Db	2521	GATGAAAACTCCACTGTTTACATAGATGCAGGCGCTGGTGATCTGTGTACCCGGGTAGAA	2580
Qy	861	SerGlyClyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAla	880
Db	2581	AGTGGAGGCTTAGTGACGCTTCAACAGGCACAGAAGTCAGATGTGCTGATCATATTCGT	2640
Qy	881	AsnGlyProLysArgSerAspAlaAlaGlnAlaValIysLysMetArgIleGluGluIle	900
Db	2641	ACCGGGCCAAAGAGAAGTGAATCCAGCTCAGGGCGGTGAAGAATGAGGATCGAGGNAATA	2700
Qy	901	GluAspAspAsnGluGlnMetIleGluAsp	911
Db	2701	GAAGATGACCAATAATCAGGAAATGATCCAGGAT	2733

US-10-310-154-113
/ Sequence 113, Application US/10310154
/ Publication No. US20030233670A1
/ GENERAL INFORMATION:
/ APPLICANT: Edgerton, Michael D
/ APPLICANT: Chomet, Paul S.
/ APPLICANT: Adams, Thomas H
/ APPLICANT: Ruff, Thomas G.
/ APPLICANT: Agarwal, Ameeta K.
/ APPLICANT: Andrews, Jeffrey E.
/ APPLICANT: Ball, James A.
/ APPLICANT: Banu, G.
/ APPLICANT: Bell, Erin
/ APPLICANT: Boddupalli, Raghava
/ APPLICANT: Deikman, Jill
/ APPLICANT: Deng, Molian
/ APPLICANT: Dong, Jinhao
/ APPLICANT: Duff, Stephen M.
/ APPLICANT: Galligan, Meshaan M.

APPLICANT: Hinchey, Brenda S.
 APPLICANT: Huang, Shihshieh
 APPLICANT: Johnson, G. Richard
 APPLICANT: Jung, Vincent
 APPLICANT: Kretzmer, Keith A.
 APPLICANT: Laccetti, Lucille B.
 APPLICANT: Lai, Chao-Qiang
 APPLICANT: Lee, Gary
 APPLICANT: Lin, Jie-Yi
 APPLICANT: Liu, Jingdong
 APPLICANT: Lu, Bin
 APPLICANT: Luethy, Michael M.
 APPLICANT: Madson, Adrian
 APPLICANT: Madson, Linda L.
 APPLICANT: Malloy, Kathleen A.
 APPLICANT: McKiel, Christine L.
 APPLICANT: Miller, Philip W.
 APPLICANT: padnavathi, Manchikanti
 APPLICANT: Parnell, Laurence D.
 APPLICANT: Start, William G.
 APPLICANT: Tennessee, Dan
 APPLICANT: Vidya, K.R.
 APPLICANT: Wang, Haiyun
 APPLICANT: Xin, Zhanguo
 APPLICANT: Xu, Nanfei
 APPLICANT: Yang, Chunzhi
 APPLICANT: Zeng, Xiaoping
 APPLICANT: Zhang, Qiang
 APPLICANT: Zhao, Yajuan
 APPLICANT: Zhou, Li

TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants

FILE REFERENCE: 38-15(52796)B
 CURRENT APPLICATION NUMBER: US/10/310,154
 CURRENT FILING DATE: 2002-12-04
 PRIOR APPLICATION NUMBER: 60/337,358
 PRIOR FILING DATE: 2001-12-04
 NUMBER OF SEQ ID NOS: 736
 SEQ ID NO 113
 LENGTH: 3942
 TYPE: DNA
 ORGANISM: Zea mays
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (436)..(1773), (1878)..(2159), (2281)..(2621), (2711)..(2990), (3079)..(3276),
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (1)..(3942)
 OTHER INFORMATION: unsure at all n locations
 10-0-310-154-113

Alignment Scores:
 Seq ID No.: 2,35e-291 Length: 3942
 Percent Similarity: 78.26% Matches: 770
 Local Similarity: 71.23% Conservations: 76
 Mismatches: 60
 Indels: 177
 Gaps: 11

09-812-350-17 (1-911) x US-10-310-154-113 (1-3942)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 436 ATGAATCCGACAACTTCAACCAAGACCAAGAGCGGATCGTGGGGGCGCAGAGATT 495
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 496 GCGGTGGAGGCGCCACGCGAGCTCACGCGCGTGCACCTGGCCGCGAGTGTGGTGGC 555
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAla---GlyGlyGluAsnAlaAla 59
 556 GACNAGGGCGGCATCTTGGCGGAGCCCATCACGGGGCGCTGCGGGGCGCGAGCGGCC 615

60 ---GlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 616 GGAGGACTGTTGAGCGCGTGTGAACAACTCGCTCAAGAGCTGCGTTCGAGTCCCG 675
 79 ProProAspAspIleProAlaSerSerLeuIleLysValIleArgArgAlaGlnAla 98
 676 CCGCCGAGCTCCGTTCCGCGCTTACGCGCTGATCAAGGTCTATCGCGCGGCGAGTCC 735
 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 736 GCGCAGAGAAACGCGGAGCTCGCACTTCGCGCGTCCAGCTGCTGCTCGCGCTGCTC 795
 119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138
 796 GAGGACTCGCAGATCTCCGACTCGCTCAAGAGGCGCGCTGTCGCGCGCGGCGTGGC 855
 139 SerGluValGluLysLeuArgGlyLysGluGlyLysLysValGluSerAlaSerGlyAsp 158
 856 GCGAGCTTGAAGAGCTCCGCGCGGAGGCGCGCGCTGAGTCCGCGCTCGCGGAGT 915
 159 ThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeu 178
 916 ACCAACTTCAGCGCTCAAGACATACGCGCGGAGCTGCTCGACGAGCGCGGAGT 975
 179 AspProValIleGlyAspGluGluIleArgArgValValArgIleLeuSerArgArg 198
 976 GACCCCGTTCATCGCGCGGAGGAGATCCGCGCGCTGCTGCGCATTTCTCTCGCGCGC 1035
 199 ThrLysAsnAsnProValLeuIleGlyProGlyValGlyLysThrAlaValValGlu 218
 1036 ACCAAAGATAACCCCGTCTCATCGCGAGGCGCGCGTTCGCAAGACGCGCTCGTGAG 1095
 219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeu 238
 1096 GGCCTCGCGCAGCGATGTTTCGCGCGAGTGCCTCCAGTAACTCTCTCGAGCTCGCCTC 1155
 239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGlu 258
 1156 ATCGCGCTCGACATCGCGCGCTCTCGCGCGGCGCAAGTACCGCGCGAGTTCGAGGAG 1215
 259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIle 278
 1216 CCGCTCAAGCGCGTCTCAAGAGAGTGGAGAGGCGCGAGGAGGAGTCAATCTCTTCATC 1275
 279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsn 298
 1276 GACGAGATACCTCTCTCGCGCGGCGAGAGCGAGGTTTCATGCGCGCGCCAC 1335
 299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
 1336 CTGTTCAAGCCAAATCTCGCGAGGAGCAGCTCAGGTGATCGCGCGCCACCACTGAG 1395
 319 GluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnValTyr 338
 1396 GAGTACCGCAAGTACGTGGAGAGAGCAGCGCTTCGAGCGCGGTTCCAGCAGGTGTT 1455
 339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyr 358
 1456 GTCCGCGAGCGCGGCTCCCGACACCGTCAAGCTTCGAGGGGCGCTCAGGAGAGTAC 1515
 359 GluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378
 1516 GAGGGGCACTTGGCTGAGGATCCAGGACCGCGCTCGTGGTTCGCGCGCACAGCTCTCC 1575
 379 AlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
 1576 GCGAGGTACATCATGGTTCGCGCATCTGCTCACAAGACATAGACCTGGTGGAGAGGCC 1635
 399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArg 418
 1636 TCGGCATGTGAGGTCAGCTCGACAGCCGCGGAGAGATGATTAACCTGGAGAGG 1695
 419 LysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAspLysAlaSer 438

1696 AAGAGATCCAGCTTGGAGTTCAGCTCCAGCGCTCGAGAGGAGGACAGGCGCAGC 1755
439 LysAlaArgLeuLeuVal----- 445
1756 AAGCGCGCTGATTGAGGT-CCGTGCTCGAGCACTGAATTTCTCAACAAAGTCCTC 1814
445 ----- 445
1815 TTGTCTGATGTTCTGTTTGAATTTGATTAACCTGCTTAACCTGCTGCTGCTT 1874
446 -----ArgLysGluLeuAspLeuArgAspLysLeuGlnProLeuThrMetLysTyr 463
1875 CAGGTCAAGGAAGATTGACGATCTGAGGCAAGCTGACGCCCTGACCATGAAGTAC 1934
464 ArgLysGluLeuValArgLeuLeuValArgLeuLeuValArgLeuLeuValArg 483
1935 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1994
484 MetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaAlaAspLeuArg 503
1995 CAGTTCAACCTCGCAGGAGCGCGCGCGGATGACCTGGCCCTGCTGGCCGACCTCAAG 2054
504 TyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThrSerGlu 523
2055 TACGGCGCCCTCCAGGAATCAGCGCTGCTATCTCAAGCTGGAG---AGCGAAACAGGG 2111
524 GluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu----- 539
2112 GAGAACCTGATGCTCACCAGAACCGTGGCGCCCTGAACAAATTCAGAGGTATGTTATTAT 2171
539 ----- 539
2172 TCTTGTTCACGCTCACAAAATTTTTCAGAGCAAGTGCAGATTTTGGCGATGCTCTCCTA 2231
540 -----ValValSerAr 543
2232 GTAGAGTAGTCTGCAGCGCTGCTGAAATGCTGTTCTGCTATGGCGAGGTGGTGAAGCG 2291
543 GTPThrGlyIleProValThrArgLeuGlyGlnAsnGlyLysGluArgLeuIleGlyLe 563
2292 TTGACGGGTATTCAGTGACCCGCTGGCCAGAACGACAGGAGGAGGCTGTTGGCT 2351
563 uAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerG1 583
2352 GGCTGACAGGCTTACCAGAGGCTGCTGGCCAGACAGAGGCTGTGAGCGCTGCAGAGA 2411
583 uAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLe 603
2412 GCGGTGCTGAGGTGAGGCGCGCTTGGCAGGCCACACAGCCCTGCTGCTGCTTCT 2471
603 uPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLe 623
2472 TTCTCTGGTCCGACTGGCGTGGGAAACTGAGTGGCCAGGCCCTAGCCGACAGCT 2531
623 uPheAspAspGluAsnLeuValArgIleAspMetSerGluTyrMetGluGlnHisSe 643
2532 GTTCACACGAGAACTTCTTGTCCGATCGACATGTCGAGTACATGAGGAGGAGGAGGAG 2591
643 rValSerArgLeuIleGlyAlaProProGly----- 653
2592 GGTGTCGCCCTCATCGGAGCACCACTGG- GTAAGTAGCAGAAAATGCATGCTTTCG 2650
653 ----- 653
2651 TATTTTAACTCGAATTCGATGAACTCTTGTCTGACAGTGGCGGAGCTGTGACAG 2710
654 -TyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValaArgArgProTyrCy 673
2711 CTACGTGGCGCATGAAGAGGTGGGAGCTGACTGAACAAAGTGAGGAGGAGGAGGAGGAG 2770
673 sValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeuGlu 693

2771 CGTGATCTCTTCGAGGAGCTCGAAGGCCCTATGTCGCGTGTCAACACCCCTGCTCCA 2830
693 nValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAnSe 713
2831 GGTCTCTCGAGCGAGGCTGCGATGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2890
713 rValIleLeuMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLy 733
2891 CGTGATCATCATGACATCGAACCTCGCGCCGAGGAGCTCTCGCTGGGATGCTGGGCAA 2950
733 sValThrMetGluValAlaArgAspCysValMet-Arg----- 745
2951 GAATCCATGAAGTCTGCTCGCATCTGCTGATGCTGATGCTGATGCTGATGCTGATGCT 3010
745 ----- 745
3011 CTGAATCGTCTGATGACATTTCTCTCTGCGAGCGTGAATGCTCAATCGGATCGTTATTGCT 3070
746 ----GluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAspGluIleValV 764
3071 GTGTGACAGGTGAGGAGGACCTTCGCGCTGAGTGTGAACCTCTCGACGAGATCGTGA 3130
764 alPheAspProLeuSerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspV 784
3131 TCTTCGATCTCTGTCCTCCACGAGCAGCTGAGGAGGCTGCTCGCTTCAGATGAGGATG 3190
784 alAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspT 804
3191 TGGCGTCTCTCTTTCGGAAGGGGATGCTCTGCTGTGACCGAGCGCGCATTTGACCA 3250
804 YrIleLeuAlaGluSerTyrAspPro----- 812
3251 TCATCTTGTCTCTCTTACGATCCGCTATGTGACCATCCATGATTTGATTCATCTGAAT 3310
812 ----- 812
3311 TCGTGGTGACACTGATGCTGCTGCTCTTATCTTTTGTGTGGCTTCAACACAG 3370
813 ValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeuSer 832
3371 GTGTATGGCGCGCGCCCATCAGAGGTGATGATGAGAGGAGGAGGAGGAGGAGGAGGAG 3430
833 LysMetValValArgGluGluIleAspGluAsnSerThrValTyrIleAspAlaGlyAla 852
3431 AAGATGTGATCCAGGAGGAGATCGACGAGACTGACCGCTCTACATCGAGCGCGCGCC 3490
853 Gly-----AspLeuValTyrArgValGlu-----SerGlyLeuValAspAlaSerThr 869
3491 GGCAGGACAGCTGCTGCTACAGGGTGGACCGGAGCGGGCTCTGCTGACCGCTGAGAGC 3550
870 GlyLysLysSerAspValLeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAla 889
3551 GGGATGAAGTCCGACATCTGATCCAGGTCCCCACAGCTCCACAGGAGGAGCGCTGCG 3610
890 GlnAlaValLysLysMetArgIle-----GluGluIleGluAspAspAsnGluGlu 907
3611 CAGCGCTCAAGAGATGAGATCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3667

RESULT 4

US-10-369-493-43734
; Sequence 43734, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 43734

LENGTH: 2643

TYPE: DNA

ORGANISM: No. US20030233675A1toc punctiforme
110-369-493-43734

Alignment Scores:

i. No.: 4,68e-181 Length: 2643
2315.00 Matches: 467
Conservative: 172
Mismatch: 199
Indels: 28
Gaps: 11

us-09-812-350-17 (1-911) x us-10-369-493-43734 (1-2643)

2 AsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAla 21
34 AATCCGAGCAATTTACGAAAGACCTGGAGACCTTGTTCGTACTCCCGAAATGGCC 93
22 ValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSerAsp 41
94 AAACAGTTTCAGCATCAGCAGATTGAGAGCGAACATTTGATGCTAGCGCTACTG--GAA 150
42 ProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGluLeuAlaAlaGlnSer 61
151 CAGGAGGATC---GCCAGTCTTATTTTCAACAAGCCGGGTAAATGTTCAAAAACCTA 207
62 AlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProAsp 81
208 CACGAGCGCACTATCGACTTCATCAACCGTCAGCCCAAGTATCGGGGGCTAGCAGTGT 267
82 AspIleProAlaSerSerLeuIleLysValIleArgAlaGlnAlaAlaGlnLys 101
268 TCAGTATACATCGACACATTTAGAGAGGCTGCTCGATCGCGCGCAACAATACCGCAA 327
102 SerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAspSer 121
328 GAGTTTGGTGAATATATTTCTTGAACATTTAATATCTTCCTTTGCCAAGACGAT 387
122 GlnIle---ArgAspLeuLeuValGlyValAlaThrAlaArgValLysSerGlu 140
388 CGCTTTGGTAAAGGTTGTTCAAGATTGGATCGATGGAATAAATCTCCCAACATC 447
141 ValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsn 160
448 ATTCAACAGATTCCGAGGAGTCAA-----AAAGTGACAGATCAAAACCGGAAGTTAAA 501
161 PheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGlnAla-----GlyLysLeu 178
502 TATGACCGGTAGAAATAACGGCGGATTTAACCAATGGCAGATGAGGCGATACCT 561
179 AspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeuSerArgArg 198
562 GACCCAGTGTGGCAGATGAAGAAATTCGCGCAGCAGATTCAATCTTCTCTCGCG 621
199 ThrIleAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGlu 218
622 ACTAAATAATACCCCGTGTGATTTGGTGAACCCGGTGTGCGTAAACCGCAATTTAGAA 681
219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeu----- 233
682 GGATTAGCGAGCGCATTTGTCAGTGGTATGTCCTCCGGAATCATTTGGCGGTGGCATAGCC 741
234 -----ThrAspValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLys 251
742 GGCCGTAGGATCGCAACTAATAGCTTTAGACATGGGTGGTAAATTTGCCGGAGCCAA 801
252 TyrArgGlyGluPheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGlu 271
|||||

802 TACCGGGGAGAAATTTGAAGAACGCTGAAAGCTGCTTTAAAGAAATCCAAAGACACAG 861
272 GlyLysValIleLeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGlu 291
862 GGACAAATCGTCTTGTTCATTGACGAATTCACACTAGTTGGTGGGGTGCACACGCAA 921
292 GlySerMetAspAlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCys 311
922 GGATCGATGATGCTAGCAACTTGTCAAGCGATGCTGCTCGCGGGGAACTGCGCTGT 981
312 IleGlyAlaThrThrLeuGluGluTyrglyTyrglyValGluLysAspAlaAlaPheGlu 331
982 ATTGTGCGCACCACTAGATGAATACCGCAAGTACATTGAAAAAGATGCGGCTTGGAA 1041
332 ArgArgPheGlnGlnValTyrglyValAlaGluProSerValProAspThrIleSerIleLeu 351
1042 CGTCGCTTTTACAGAGGTATATGTCGATCAGCCAGTGTGGAAGATACCATCTCTCAATTCTG 1101
352 ArgGlyLeuLysGluLysTyrglyHisHisGlyValArgIleGluAspArgAlaLeu 371
1102 CCGGTTTGAAGAGCGCTACGAGTTACACACGCGGTGAAGATTTCTGTAGTCGCTTA 1161
372 IleAsnAlaAlaGlnLeuSerAlaArgTyrglyThrGlyArgHisLeuProAspLysAla 391
1162 GTTGGCGGAGCTACTCTGTCGAGATATATTAGCGATCGCTTCTTACCCGACAAAGCT 1221
392 IleAspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGlu 411
1222 ATTGATTTGGTGGATGAAGCTCGCGCTAAACATAAATATGGAATTTACTTCCAAACCGAA 1281
412 GluIleAspAsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGlu 431
1282 GAATTAGACGAATCGATCGCAAAATTTGCACTGGAATGGAGCGGCTGTCTGCTGAA 1341
432 ArgGluLysAspLysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAsp 451
1342 AAAGAAACAGACAGTGTCTCCAGAGAGCGTTTAGAACGGTGGAGAGAGAACTTGCAGAG 1401
452 LeuArgAspLysLeuGlnProLeuThrMetLysTyrglyLysGluLysGluArgIleAsp 471
1402 TTAAGAAGACCGCAAGATGCCCCCTCAATGCTCAATGGCAAGCAGAGCAAAATCATCGAT 1461
472 GluIleArgArgLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGlu 491
1462 CCATTCGCCAAATTAGCAAGAGATTGAGCGTGTCAATGTAGAAATTCAGCAAGCCGAA 1521
492 ArgArgTyrglyAspLeuAlaArgAlaAlaAspLeuArgTyrgly----- 505
1522 CGCGATTACGACCTCAACCGGCGCAGCAGAAATGAAATACAGCAAACTCAGTGAGCTGCAA 1581
506 ---AlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGlu 524
1582 CGAACACTAAAGAAGCGCAAGCGCAGCTAGCCCAATTCAG---ACTAGCGGTAAATCT 1638
525 AsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyr 544
1639 -----CTGTTCGCGAAGAGTCAACCGAGCTGACATTCGCGAAATTTCTCCAGTGG 1692
545 ThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAla 564
1693 ACGGTATTCGGTGAGCAGACTGTTGAATCGGAAATGCAAAACTCTTGCATCTCGAA 1752
565 AspArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAla 584
1753 GAAGAGCTACATAAACCGTGTGATTTGGTCAGACGAGCTGTGAAGAGCTGTTCGCGATGCA 1812
585 IleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPhe 604
1813 ATTCACGTTCCCGTGTGGTGGCAGACCCGAAATCGTCCATTCGTAGCTTTATTTTC 1872
605 LeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPhe 624
1873 TTAGCCCGCAGCCGGGTGGGAAACAGAACTGCAAAAGCTTTAGCGGAATATCTGTTT 1932

ULT 5
 10-369-493-26449
 Sequence 26449, Application US/10369493
 Publication No. US2003023675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianteng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 26449
 LENGTH: 2619
 TYPE: DNA
 ORGANISM: Synechocystis sp.
 10-369-493-26449

Db 889 GCGGGCAACTATTTGAAACCCATGTTAGCCCGGGTGCTTTGCGTTGTATCGGGGCCACC 948

316 ThrLeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgPheGln 335
 949 ACTTTAGATGAATATCCAAATATATCGAAAGAGATCGGCTTTGAGACAGCGTTCCAG 1008
 336 GlnValTyrValAlaGluProSerValProAspThrLysSerLysLeuArgGlyLeuLys 355
 1009 GAAGTTTTAGTGGATGAACCAATGTTATAGATACCAATTTCCATTTCTCCGGGATTAATA 1068
 356 GluLysTyrGluGlyHisHisGlyValArgLysGlnAspArgAlaLeuLysLeuAlaAla 375
 1069 GAAGCTATGAAGTACACCGCGGTAATAATTCGCCATAGTGCCTGTGAGCGGCGCC 1128
 376 GlnLeuSerAlaArgTyrLysLeuThrGlyArgHisLeuProAspLysAlaLysLeuVal 395
 1129 ATGTTGTCCCAATCTTACATCAGTATCGTTTCTGCGCGATTAAGCTATTGATTTAGTA 1188
 396 AspGluAlaCysAlaAenValArgValGlnLeuAspSerGlnProGluGluLysLeuAspAen 415
 1189 GACGAAGCAGCGGCAAAATTAATAATGGAATCACTCCAAACGAGGGAATTAGATGAA 1248
 416 LeuGluArgLysArgMetGlnLeuGluLysLeuHisAlaLeuGluArgGluLysAsp 435
 1249 GTTGACCGGAAATTCCTCAATAGAAATCGAGCGTTATCTTTACACCGGAAATGAT 1308
 436 LysAlaSerLysAlaArgLeuLysLeuValArgLysGluLeuAspAspLeuArgAspLys 455
 1309 TCTGCTTCCAGGAGCGGTAGAAAATTCGAGAAAGATGCTGATTTTAAAGAAAGAA 1368
 456 LeuGluProLeuThrMetLysTyrArgLysGluLysGluArgLysLeuArgArg 475
 1369 CAGTCTAACTTAATGCGCAATGGCAGTCGGAATAACCGGTTATTCATCAAAATTCGTACT 1428
 476 LeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAsp 495
 1429 GTTAAGGAACCATCGACCGAGTGAACCTAGAAATTCACAGCGCCCAACCGGATTCAGAC 1488
 496 LeuAlaArgAlaAlaAspLeuArgTyrGlyValAlaLeuGlnGluValGluSerAlaLeuAla 515
 1489 TACAATAAAGCAGCGAGTTACAGTATGCAATTAACCTGATTTACAGCGGCAAGTGAA 1548
 516 GlnLeuGluGlyThrSerSerGluGluAenVal-----MetLeuThrGluAen 531
 1549 GCTTTGAAACCAATTTGGCGGAGCAACAACTCTGGCAATTCCTCTTACCGGAAGAA 1608
 532 ValGlyProGluHisLeuAlaGluValValSerArgTyrThrGlyLeuProValThrArg 551
 1609 GTTTAGAGTCTGACATTCGTGAATTAATCTCGAATGCGCGCATTCCTCATCAGTAA 1668
 552 LeuGlyGlnAsnGluLysGluArgLeuLysLeuAlaAspArgLeuHisLysArgVal 571
 1669 TTGGTGAATCGGAAAGAAAACCTGCTCACTTGAAGATGAACCTACACAGCGGAGTG 1728
 572 ValGlyGlnAsnGlnAlaValAenAlaValSerGluAlaLeuArgSerArgAlaGly 591
 1729 ATTGGTCAGATGACGGTAACCGCGTAGCGAAGCACTTCAACGCTCCCGAGCTGT 1788
 592 LeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGly 611
 1789 CTTTCGATCTTAATCGTCCACCGCTAGCTTTATTTCTGGGCCCCACAGGGGTGCGG 1848
 612 LysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGlnLeuVal 631
 1849 AAAAATGAGTTAGCGAGGCTTTGGCGGAAAAATTTATTCACACGGAAGACCTCTGGTG 1908
 632 ArgLeuAspMetSerGluTyrMetGluGlnHisSerValSerArgLeuLysLeuAlaPro 651
 1909 CGAATGATATGCTGAATATATGGAATAACAGCTGTTTCCCGTTTATATGGGGGCCCT 1968
 652 ProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArgPro 671
 1969 CCGGGCTATGCGGCTATGAAGAGGGGAGCAATTCAGCGAAGCAATTCGCGCGCGGCC 2028

672 TyrCysValLeuLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeu 691
 2029 TATTCGGTCATTTCTTTTGACGAGATGAAAGAGCCATGGGATGTGTTAACGTCATG 2088
 692 LeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArg 711
 2089 CTCCAATCTCGATGATGCGCGTTTAAACGATGCCAAGGCCATGTGTGGATCTCAA 2148
 712 AenSerValLeuLeuMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThr 731
 2149 AATAGCATTTATCATTTATGACCAAGTAACTGGGCTCCCAATACATTTTGGATGTGGCGGG 2208
 732 GlyLysValThrMetGluValAlaAlaArgAspCysValMetArgGluValArgLysHisPhe 751
 2209 GATGATAGTCGTTATGAAGAAATGCGGAGCCGAGTTATGGATGTAATCGGGAAACCTTC 2268
 752 ArgProGluLeuLeuAsnArgLeuAspGluLeuValPheAspProLeuSerHisAsp 771
 2269 CGCCCAAGATTTCTCAATCGGCTGGATGAAACGATTTATTTCCATGGCTTACAAAAATCC 2328
 772 GlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGluArg 791
 2329 GAGTTACGATTCATTTGTCCTCAATTCAAATTCAGCTTTGGCTACCCGTTTGGAGGAAACA 2388
 792 GlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrLysLeuAlaGluSerTyrAsp 811
 2389 AAATTAACCTTTGAAGCTTAACGAGTAAGCCCTAGATTTCTGGCTGCGCTATGAC 2448
 812 ProValTyrGlyAlaArgProLysLeuArgArgTrpMetGluLysLysValValThrGluLeu 831
 2449 CCGCTTTATGGGCGCGGACCTTTAAACAGCGCGTCCAAAAATACCTAGAAACCGCGATC 2508
 832 SerLysMetValValArgGluGluLysLeuAspGluAenSerThrValTyrLysAlaGly 851
 2509 GCCAAGGAATTTTACGGGGGATTCACAACTGGTGAGACCATTTGGTGGATGAAC 2568
 852 AlaGlyAspLeuValTyrArgValGluSerGlyGlyLeuVal 865
 2569 GACGAACGCTTCAGTTTACCAGTTTAAAGGGGGGATTTAGTC 2610

RESULT 6

US-10-047-260-37
 ; Sequence 37, Application US/10047260
 ; Publication No. US20020164706A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Huang, Lina
 ; APPLICANT: McClusky, Michael
 ; APPLICANT: Lakoska, Robert
 ; TITLE OF INVENTION: HIGH LEVEL PROMOTERS FROM CYANOBACTERIA
 ; FILE REFERENCE: CL1715 US NA
 ; CURRENT APPLICATION NUMBER: US/10/047,260
 ; CURRENT FILING DATE: 2002-01-15
 ; PRIOR APPLICATION NUMBER: 60/264,925
 ; PRIOR FILING DATE: 2001-01-30
 ; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 37
 ; LENGTH: 2619
 ; TYPE: DNA
 ; ORGANISM: Synechocystis sp. strain PCC6803
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2619)
 US-10-047-260-37

Alignment Scores:
 Pred. No.: 7,74e-179 Length: 2619
 Score: 2288.00 Matches: 460
 Percent Similarity: 71.74% Conservative: 167
 Best Local Similarity: 52.63% Mismatches: 229
 Query Match: 50.29% Indels: 18
 DB: 14 Gaps: 7

09-812-350-17 (1-911) x US-10-047-260-37 (1-2619)

2 AsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAla 21
 13 GATCCTTAATAATTTACGGAGAAAGCTTGGAGGCGCATCGCAAAACACCGAGATGCT 72
 22 ValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuLeuSerAsp 41
 73 AAACAGCATCGCAACAGCAAAATTTAGACGGAACACCTACTCAGTGCCTACTA--GAA 129
 42 ProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGlnSer 61
 130 CAANAATGGTTCGCCACCGCATCTTTAATAATAGCTGGGGCG-----AGC 174
 62 AlaGluArgVal-----IleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
 175 ATTCGCCGAGTTAACGATCAAGTTAATAGCTTTATGCCCCAACAGCCAAATTAAGTAAT 234
 79 ProProAspIleProAlaSerSerSerLeuIleLysValIleArgArgAlaGlnAla 98
 235 CCGAGTGAATCGATTTATTAGCCCGCAGTCTCGATAAATTTGTGGACAATCGGAAATA 294
 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
 295 GCCAGTCTAATATGAGACGACGACTATATTTTCATCGACACTTGATGGCGGCTTACGGC 354
 119 GluAspSerGlnIle---ArgAspLeuAsnGluValGlyValAlaThrAlaArgVal 137
 355 CAAGATGACCGCTGGGCAAAATTTATATCGAGAAATTTGGCTAACAGAAAATAAGTTG 414
 138 LysSerGluValGluLysLeuArgGlyLysGluGlyLysValGlnSerAlaSerGly 157
 415 GCAGAAATTTATCAAGCAAAATTTAGGAACCCAA-----AAAGTGACCGATCAAAATCCA 468
 158 AspThrAsnPheGlnAlaLeuLysThrTyArgAspLeuValGluGlnAla----- 175
 469 GAGGGCAATACGATCCCTTGAAAATATGGCGAGATTTAACGGAATTAGCCCGGAA 528
 176 GlyLysLeuAspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeu 195
 529 GGTAAACTAGATCTGTCTATGGCGGATGAAGAAGTGGCGGCACCATTCAGATCCTT 588
 196 SerArgArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAla 215
 589 TCCCGCGGCACAAAAAATACCTGTGTTAATTTGGGGAACCGGGGGTGGTAAACCGCG 648
 216 ValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAsp 235
 649 ATCCCGGAGGTTTAGCCCAAGAATTTATACCATGACGTACCGGAATCATTTGGGGAT 708
 236 ValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyArgGlyGlu 255
 709 CGCAAACTAATTTCCCTCGATATGGGGCGGTAAATTTGGCGGGCAAAATACCGGGGGAA 768
 256 PheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIle 275
 769 TTTGAGAAAGACTTAAGCGGTACTTTAAGAAGTTTACCGAGCCGAGGGGCAATATT 828
 276 LeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAsp 295
 829 CTCTTTATTGACGAATTCATACCTGTGTGGCGCTGGGGCCACCCAGAGAGCATGAT 888
 296 AlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThr 315
 889 CGGGGCACCTTATTGAACCCATGTTAGCCCGGGGTGCTTTGCGTGTATCGGGGCCACC 948
 316 ThrLeuGluGluTyArgLysTyArgValGluLysAspAlaAlaPheGluArgArgPheGln 335
 949 ACTTTAGATGAATATCGCAAAATATATGAAAAGATCGGCTTTTGGAAACGACGTTTCAG 1008
 336 GlnValTyValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLys 355
 1009 GAAGTTTATAGTGAATGAACCAATGATTATAGATACCAATTTCCATTTCCCGGGGATTA 1068

356 GluLysTyArgGluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAla 375
 1069 GAACGCTATGAAGTACACCGCGTAAATAATTCGGATAGTGCCTCGGTAGCGCGCGCC 1128
 376 GlnLeuSerAlaArgTyIleThrGlyArgHisLeuProAspLysAlaIleAspLeuVal 395
 1129 ATGTTGTCCCAATCGTATCATCAGTGATCGTTTCTGCGGATAAAGCTATTGATTAGTA 1188
 396 AspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsn 415
 1189 GACGAGCAGCGGCAAAATTAATAATGAAATCACTCCAAACCCAGAGGAATTAGATGAA 1248
 416 LeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLysAsp 435
 1249 GTTGACCGGAAATTCCTCAACTAGAAATCGAGCGTTTATCTTTACACCGGAAATGAT 1308
 436 LysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspLeuArgAspLys 455
 1309 TCTGCTTCCAGGAGCGCTAGAAAATTTGGAGAAAGAGTTGCTGATTTTAAAGAAGAA 1368
 456 LeuGlnProLeuThrMetLysTyArgLysGluLysGluArgIleAspGluIleArgArg 475
 1369 CAGTCTAAACTTAATGCCCAATGCGCAGTCGGAATAACGGTTATTGATCAAAATCGTACT 1428
 476 LeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyArg 495
 1429 GTTAAAGGAACCATCGACAGGTGAACCTAGAAATTTCAACAGGCCCAACCGGATTACGAC 1488
 496 LeuAlaArgAlaAlaAspLeuArgTyArgLysAlaIleGlnGluValGluSerAlaIleAla 515
 1489 TACAATAAGCAGCGGAGTTACAGTATGCAATTAATTAAGTTTACAGCGGCAAGTGA 1548
 516 GlnLeuGluGlyThrSerSerGluGluAsnVal-----MetLeuThrGluAsn 531
 1549 GCTTTGGAACCCCAATTTGGCGGAGCAACAACTCTGGCAAAATCCCTCTTACGGGAAGAA 1608
 532 ValGlyProGluHisIleAlaGluValValSerArgTyThrGlyIleProValThrArg 551
 1609 GTTTTAGAGTCTGACATGCTGAAATTTATCTCGAAATGACCGGCATTCCTCATCAGTAA 1668
 552 LeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgVal 571
 1669 TTGTTGGAATTCGGAAGAAAACAACTGCTCACTTGGAAAGATGAATACACAGCCGAGTG 1728
 572 ValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGly 591
 1729 ATTGTCAGATGACGCGTTAACCGCGGTAGCCGAGGCCATTCACGCTCCCGAGCTGT 1788
 592 LeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGly 611
 1789 CTTTCCGATCCTAATCGTCCACCGCTAGCTTTATTTTCTGGGCCCCACAGGGTCCGG 1848
 612 LysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuVal 631
 1849 AAACCTGAGTTAGCAAGGCTTTGGCGAAAATTTATTCGACCGAAGAACCCCTGGTG 1908
 632 ArgIleAspMetSerGluTyMetGluGlnHisSerValSerArgLeuIleGlyAlaPro 651
 1909 CGGATTGATATGCTGAATATATGGAACAAACACGCTGTTTCCGTTTAAATGGGGCCCT 1968
 652 ProGlyTyThrValGlyHisGluGluGlyGlnLeuThrGluAlaValArgArgArgPro 671
 1969 CCGGGCTATGTGGCTATGAGAAGGGGCAATTTGACGGAAGCAATTCGCGCGCGGCC 2028
 672 TyrcysValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeu 691
 2029 TATTCGCTCATCTTTTTCAGGAGATGAAAAGGCCCATTTGGGGATGTGTTTAACTCATG 2088
 692 LeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArg 711
 2089 CTCCAATCTCGATGATGCGCTTTTACCGATGCCAAGGCCATGTGTGTGACTTCAA 2148

APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHIYUKI
 APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 CURRENT FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109
 EQ ID NO 7220
 LENGTH: 2637

SEQUENCE: 2,436-178 Length: 2637
 Matches: 460
 Percent Similarity: 71.35%
 Conserved: 160
 Mismatches: 231
 Indels: 18
 Gaps: 8

SEQUENCE: 10-156-761-7220 (1-2637)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
 1 ATGGACATGAACCGCTCACCAGAGTCCAGGAGGCTCCAGGAGGCGCCAGCCGCG 60
 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
 61 GCCGCGCGCATGGGGGACACACGAGTGCAGCGGGAACACCTGTCTGCTCGACTTCTCGAT 120
 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
 121 CAGGAGGACGCTGTATCCCGTGTGTCACAGCGCGGACCGGACCGGACCGGAG 174
 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProPro--- 79
 175 GAATGCGCGCGCGCGTGGCGGAGAACTCTCCACCGCGCGGAGGCGCGCGCGCG 234
 80 -----ProAspAspIleProAlaSerSerLeuIleLysValIleArgArgAlaGln 97
 235 GCGGACCCCGCGCGAGTCTTGTCTCACCAGCGCTCGCGGCTGCTCGACGCGCGCGAG 294
 98 AlaAlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeu 117
 295 CGGAGGCGCAACCGCTCAGGACAGTACGTGCTGCGTGAGACCTCTGCTGCGCGCTG 354
 118 LeuGluAspSerGlnIleArgAsp-----LeuLeuAsnGluValGlyValAlaThr 134
 355 GCTGAGGAGAGCTCTTCGACCGCGCGCGGCTACTGTCTCAACAGCGCGCGCATCACCAGG 414
 135 AlaArgValLysSerGluValGluLysLeuArgGlyLysGluGlyLysValGluSer 154
 415 GACTGTTCTGAGCGCGCTCACCAGGTCGCGGCAACACG-----CGGGTCACTCTCC 468
 155 AlaSerGlyAspThrAsnPheGlnAlaLeuLysThrTyrglyArgAspLeuValGluGln 174
 469 GCCAACCCGAGTGGCTACGAGGCTCTGGAGAGTACCGCGCGGACCTTGTCTCTCGAG 528
 175 Ala-----GlyLysLeuAspProValIleGlyArgAspGluGluIleArgArgValVal 192
 529 GCTCGTTCGCGCGCGCTGGAGCCCGGTCTATCGCGCGGACGCGGAGATCCGCGCGGTGACC 588

193 ArgIleLeuSerArgArgThrLysAsnProValLeuIleGlyGluProGlyValGly 212
 589 CAGATCTCTGCGCGCAAGACCAAGAACACCCCTCTCATCGCGGACCCCGCGCTCGGC 648
 213 LysThrAlaValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSer 232
 649 AAGACCGCCATCGTCGAGGCGCTGGCGCAGCGCATGTTTCGCGCGGACGTCCTCCGAGGGC 708
 233 LeuThrAspValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyr 252
 709 CTGCGCGCAAGACGCGTTCGCGCTCGACATGCGTCTCTGCTGCGCGCGCGCAAGTAC 768
 253 ArgGlyGluPheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGly 272
 769 CGCGGGAGTTCGAGGACCGCTCAGGCGCTGCTCAGCGAGGTCAAGGCCCGCGAGGGG 828
 273 LysValIleLeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGly 292
 829 CGGATCTCTCTTCTGTCGACGAACTCCACACCGCTCTAGGAGCGCGCGCGCGGAGGG 888
 293 SerMetAspAlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIle 312
 889 GCCATGACCGCGGCAACATGCTCAAGCGCATGCTCGCGCGCGCGGCAACTCCACATGATC 948
 313 GlyAlaThrThrLeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArg 332
 949 GCGCGCACCATCTCGACGAGTACCGCAGGACATCGAGAGGACCGCGCTCGAACGC 1008
 333 ArgPheGlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArg 352
 1009 CGCTTCAGCAGGCTCTGCTCGACGAGCGGAGCGTGGAGGACACCATCTCTCCATCTCTGCGC 1068
 353 GlyLeuLysGluLysTyrGluGlyHisHisGlyValArgIleGluAspArgAlaLeuIle 372
 1069 GGAATCGCGGAGCGCTCGAGGTCTTCCAGCGGCTGAAGTCCAGGACACCGCGCTGTGTC 1128
 373 AsnAlaAlaGlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIle 392
 1129 TCCGCGCGCACCTCAGCCACCGCTACATCAGGATCGGTCTCTGCGCGACAGGCCATC 1188
 393 AspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGlu 412
 1189 GACCTCTCGACGAGCGTGCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1248
 413 IleAspAsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArg 432
 1249 CTCGAGCATCACCAGCGCGCTCACCCTGAGATCGAGGAGCGCGCTGCTGCTGCTGCTGCT 1308
 433 GluLysAspLysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspLeu 452
 1309 GAGAGCGACCCCGCAGCAAGACCGCTGAGGAGCTGCGCAGGGAACCTGCGCGACCTG 1368
 453 ArgAspLysLeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGlu 472
 1369 CCGCGGAGCGCGCAACACCGCTGAGGAGCGCGCAACCGCGCGGATCCGCGCTGCTGCTGCT 1428
 473 IleArgArgLeuLysGlnLysArgGluGluMetPheSerLeuGlnGluAlaGluArg 492
 1429 GTGCGAGAACTGCGCGGAGAACTGAGGAGTCCCGCAGGCGGAGGAGGAGCGCAAGCGC 1488
 493 ArgTyrAspLeuAlaArgAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSer 512
 1489 GCCTACGACCTCAACCGCGCGCGCAACTCCGCTACGCGCTGCGCGCTGCGGAGCGCGC 1548
 513 AlaIleAlaGlnLeuGlu-----GlyThrSerSerGluGluAsnValMetLeuThr 529
 1549 CGACTGCGCGCGGAGGAGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1608
 530 GluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThrGlyIleProVal 549
 1609 GAGGTCTGTCAGGAGGAGGAGATCGCGGAGATCGCGCGCTGCGCGCTGCGCGCGCGCGC 1668
 550 ThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLys 569

```

1669 GCCGCTCCAGAGGGGCAACGAGAACTGTCGCCCTCCGAGAGATCTCCGCGGAG 1728
570 ArgValValGlyGlnAenGlnAlaValSerGluAlaLeuArgSerArg 589
1729 CGCGTCATCGCCAGGAGGAGCGTCAACTCGTCCAGGAGCCATCATCCGCGCGC 1788
590 AlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuProThrGly 609
1789 TCCGCGATCCGCGACCTCCGCGCGCCATCGGCTGTTCTCTCCGCGCCACCGCGC 1848
610 ValGlyLysThrGluLeuAlaLysAlaLeuAlaGlnLeuPheAspGluAsnLeu 629
1849 GTCGGAGAGACCGAGCTGGCAACACCTCGCCCGGACTGTTCGATCCGAGGAGAAC 1908
630 LeuValArgLysAspMetSerGluThrMetGluGlnHisSerValSerArgLeuLeGly 649
1909 ATGTCCTCGCTCGCATGAGCGAATACGAGGAGCGGACACCGCTCAGCGGCTCATGGC 1968
650 AlaProGlyGlyValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArg 669
1969 GCACCGCGCGATACGTCGCTACGAGGAGGCGCGCAGCTCACCGAGGCGGTACGCGCG 2028
670 ArgProThrCysValLeuLeuPheAspGluValGluLysAlaHisValAlaValPheAsn 689
2029 AAGCGGTACTCGGTGCTGTCGAGAGATCGAGAGCGGACACCGGTGTTCTCAAC 2088
690 ThrLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAsp 709
2089 ACCCTGCTCGAGATCTTCGAGCGGCGGCGCATCCGACCGCCCGCGCGCACCGTGCAC 2148
710 PheArgAsnSerValLeuLeuMetThrSerAsnLeuGlyAlaGluHisLeuAlaGly 729
2149 TTCGCGAACCCTGATCATCATGATGCTCAACATCGGCTCCGAGGACCTTCTCGACGCG 2208
730 LeuThr-----GlyLysValThrMetGluValAlaArgAspCysValMetArgGluVal 747
2209 GCCACCGCGGAGAGTGAGATCAAGCCCGAC---GCCGCGCCCTGCTGATGCGGAGCTG 2265
748 ArgLysHisPheArgProGluLeuLeuAsnArgLeuAspGluLeuValValPheAspPro 767
2266 CGCGGCGACATTCGCGCGGAGTTCCTCAACCGCGTCAACATCGTGTGTTCAACCG 2325
768 LeuSerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArg 787
2326 CTGGGTGAGCGGAGATCGAGCGGATGTTGGAATCGAGTTCGACGAGCTGCGGCGCG 2385
788 LeuAlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrlleLeuAla 807
2386 CTCGCGGACCGCGCATCACCGTCAACTCACCGAGCGCGCGCGGAGTGTATCGCCAC 2445
808 GluSerTyrlleAspProValTyrlleAlaArgProIleArgArgTrpMetGluLysVal 827
2446 CAGGCTACGACCGGTGTACGCGGCGCGCGCGCTGCGCGCTGATCTCCACGAGGTC 2505
828 ValThrGluLeuSerLysMetValValArgGluGluLeuAspGluAsnSerThrValTyrl 847
2506 GAGACATGTCGCGCGCGCTGTCGCGCGCGGAGCTCCAGGACGCGCGGAGTTCGCG 2565
848 IleAspAlaGlyAlaGlyAspLeuVal 856
2566 GTCGAGCGCGGAGCAGGAGAGTGGTG 2592

```

3ULT 9

Sequence 1, Application US/10156761
Publication No. US20030119018A1

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI

```

; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
; US-10-156-761-1

Alignment Scores:
Pred. No.: 4,6e-174 Length: 9025608
Score: 2282.00 Matches: 475
Percent Similarity: 68.69% Conservative: 170
Best Local Similarity: 50.59% Mismatches: 258
Query Match: 50.15% Indels: 37
DB: 15 Gaps: 12

US-09-812-350-17 (1-911) x US-10-156-761-1 (1-9025608)
QY 1 MetAenProGluLysPheThrHisLysThrAsnGlnThrIleAlaThrAlaHisGluLeu 20
Db 8634077 ATGACATGAACCGTCTCACCAGAGTCCAGGAGGCCCTCCAGGAGGCCACACCGG 8634136
QY 21 AlaValAenAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
Db 8634137 GCCGCGCGCATGCGGCGCACACCGAGGTGCGACGCGGAAACACCTGTGCTCGCACTTCTCGAT 8634196
QY 41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAenAlaAlaGln 60
Db 8634197 CAGGAGGAGGTCTGATCCCGGTTCTGCAACAGCGCCGCGACCGAG-----CCGAG 8634250
QY 61 SerAlaGluArgValIleAenGlnAlaLeuLysLysLeuProSerGlnSerProPro--- 79
Db 8634251 GAACTGCGCGCGCGGTGCGGAGGAACTCTCCACCGCGCGAGCGCGACCGCGCGG 8634310
QY 80 -----ProAspApIleProAlaSerSerSerLeuIleLysValIleArgAlaGln 97
Db 8634311 GCGGACCGCGCGCGAGGTCTTCTGTCACCGCGCGCTCCCGGCTGCTCGACGCGCGCG 8634370
QY 98 AlaAlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeu 117
Db 8634371 CGGAGGCCAACAACCGCTCAAGGAGAGTGTGCTGAGGACCTCTCTGCTGCGGCTG 8634430
QY 118 LeuGluAspSerGlnIleArgAsp-----LeuLeuAenGluValGlyValAlaThr 134
Db 8634431 GCTGAGGAGAGCTCTTCGACCGCGCGCGGTACTGCTCAACAGCGCGCATCACCGAG 8634490
QY 135 AlaArgValLysSerGluValGluLysLeuArgGlyLysGluGlyLysLysValGluSer 154
Db 8634491 GACTCGTCTCTGAGCGCGCTCACCGCGCGCGGTATCGCGCGCGGAGATCCCGCGGTGACC 8634544
QY 155 AlaSerGlyAspThrAenPheGlnAlaLeuLysThrTyrlleArgAspLeuValGluGln 174
Db 8634545 GCCAACCGCGAGTGGCTTACGAGGCTCTGGAGAGTACGCGCGCGAGCTGTGCTCTCGAG 8634604
QY 175 Ala-----GlyLysLeuAspProValIleGlyArgAspGluGluIleArgValVal 192
Db 8634605 GCTCGGTCTCGGCGCGGTGAGACCGGTCTATCGCGCGCGGAGATCCCGCGGTGACC 8634664
QY 193 ArgIleLeuSerArgThrLysAenAenProValLeuIleGlyLysProGlyValGly 212

```

863474655 CAGATCTCTAGCCGCAAGACAAGAAACAACCCCGTCTCTATCGCGCAACCCGCGCGTCCGC 86347472

213 LysThrAlaValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSer 232
863474725 AAGACCGCATCTGTGAGGGCTGGCCAGGGCATCGTTGCGCGACAGTCCCGGAGGCG 863474784

233 LeuThrAspValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyr 252
863474785 CTGCGCACAGACGGTGTTCGCCCTCGACATGGGCTCCCTGGTCGCGCGGCCCAAGTAC 86347484

253 ArgGlyGluPheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGly 272
8634845 CGCGGGAGTTCAGGAGACGCCCTCAAGCGCGTCTCAGCGAGGTCAAGCGCGCGCGAGG 8634904

273 LysValIleLeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGly 292
8634905 CGGATCTGTCTTCGTCGAGCAACTCCACACCGTCGTAGACGGCGCGCGGAAGGG 8634964

293 SerMetAspAlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGluLeuArgCysIle 312
8634965 GCCATGGACGGCGGCACATGCTCAAGCGCATGCTCGCGCGCGCAATCCACATGATC 8635024

313 GlyAlaThrThrLeuGluGlyTyrArgLysTyrValGluLysAspAlaAlaPheGluArg 332
8635025 GCGGCCACACACCTCGACGAGTACCGAAGACATCGAAGAGGACGCGCCTCGAACGC 8635084

333 ArgPheGlnGlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArg 352
8635085 CGCTTCAGCAGTCTGTGTCGACGAGCGGAGCGTGGAGGACACCATCTCCATCTCTGCCG 8635144

353 GlyLeuLysGluLysTyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIle 372
8635145 GGACTCGCGAAGCGCTCGAGGTCTTCCACGGCGTGAAGATCCAGGACACCGCGCTGTC 8635204

373 AsnAlaAlaGlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIle 392
8635205 TCGCGGCCACCTCAGCCACCGCTACATCACGATCGGTCTCTGCCGACAGGCCATC 8635264

393 AspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGlu 412
8635265 GACCTCGTCGACGAGCGTGGCGCGCTACCGAGATCGATCTGATGCGCGGGA 8635324

413 IleAspAsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArg 432
8635325 CTCGACAGATCACCGCGCGCTCACCGCGCTGGAGATCGAGGAGCGGCCCTGTCTCAAG 8635384

433 GluLysAspLysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspLeu 452
8635385 GAGCGACCCCGCAGCAAGACCGCTTGGAGAGTCTGCGCGAGGAATCGCCGACCTG 8635444

453 ArgAspLysLeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGlu 472
8635445 CGCGCGGAGCGCGACGCCAAACACGCCAGTGGAGCGCGAAGCGCGCATCGCGCG 8635504

473 IleArgArgLeuLysGlnLysArgGluGluMetPheSerLeuGlnGluAlaGluArg 492
8635505 GTGCGAGAACTCGCGCAGGAATCGAGCAGGTCCGCCACGAGCGCGGAGGAGCGCGA 8635564

493 ArgTyrAspLeuAlaArgAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSer 512
8635565 GCCTACGACCTCAACCGCGCGCGCAATCGCTACGCGCGCGCTCAGAGCTGTGAGGCG 8635624

513 AlaIleAlaGlnLeuGlu-----GlyThrSerSerGluGluAsnValMetLeuThr 529
8635625 CGACTCGCGCGGAGGAGGAGCACTGGCCGCCAAGCAAGGAGGGAACCGCGCTCTGCCG 8635684

530 GluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThrGlyIleProVal 549
8635685 GAGTGTGTCAACGAGAGGAGATCGCCGAGATCGTCGCGCGCTCGACCGCATCCCCGTC 8635744

550 ThrArgLeuGlyGluAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLys 569
8635745 CCGCGCTTCAGGAGGCGCAACGAGGAAAGCTGCGCGCTCGACGAGATCTCTCGCGAG 8635804

QY	570	ArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArg	589
Db	8635805	CGCGTCATCCGCGCAGGACGAAGCGGTCAAACCTCGTCACCGAGCCATCATCCGCGCCCGC	8635864
QY	590	AlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuGlyProThrGly	609
Db	8635865	TCGGCATCCGCGACCTCGCGGCCCATCGGCTGTTTCATCTTCCTCGCGCCACCGCGC	8635924
QY	610	ValGlyIleThrGluLeuAlaIleAlaLeuAlaGluGlnLeuPheAspArgGluAsnLeu	629
Db	8635925	GTCCGGAAGACCGAGTCGCGCAAGACCTCGCCCGGACTCTGTTCGATCCGAGGAGAAC	8635984
QY	630	LeuValArgIleAspMetSerGluThrMetGluGlnHisSerValSerArgLeuIleGly	649
Db	8635985	ATGTGTCGCCCTCGACATGAGCGAATACACGAGGCGGCACACCGCTCAGCCGGCTCATCGGC	8636044
QY	650	AlaProProGlyTyrValGlyHisGluGluGlyGlnLeuThrGluAlaValArgArg	669
Db	8636045	GCACCGCCCGATACGTCGCTACGAGAGGCGGCAGCTCACCGAGGCGGTACCGCCG	8636104
QY	670	ArgProTyrCysValIleLeuPheAspGluValGluIleAlaHisValAlaValPheAsn	689
Db	8636105	AAGCGCTACTCGTCGTCTTCACGAGATCGAAGAGCGCACACCGATGCTTCAAC	8636164
QY	690	ThrLeuLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAsp	709
Db	8636165	ACCTGTGTCAGATCTTCGACAGCGCGCATCACCGACGCCGAGGCGCACCGCTGAC	8636224
QY	710	PheArgAsnSerValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGly	729
Db	8636225	TTCGCAACACCGTGATCATCATGAGCTCCAAACATCGGCTCCGAGCACCTCTTCGACGC	8636284
QY	730	LeuThr-----GlyIleValThrMetGluValAlaAlaArgAspCysValMetArgGluVal	747
Db	8636285	GCCACGCCGAGGTGAGATCAAGCCCGAC--GCCCGCGCCCTGTGTATGGCGGAGCTG	8636341
QY	748	ArgIleHisPheArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspPro	767
Db	8636342	CGCGGGCACCTTCGCGCGGATTCCTCAACCGCGTCGACGACATGTCGTCTGTTCAAACCG	8636401
QY	768	LeuSerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArg	787
Db	8636402	CTGGGTGAGCGCGAGATCGAGCGATCTGTGAACCTGCATTCGACGAGCTGCGGCGAGCG	8636461
QY	788	LeuAlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAla	807
Db	8636462	CTCGCGAAGCCGCGATCACCGTCGAACTCACGACGCGCGCGGAGTAGTACGCCAC	8636521
QY	808	GluSerTyrAspProValTyrGlyAlaArgProIleArgArgTyrPheMetGluLysVal	827
Db	8636522	CAGGGCTACACCGGTGTACGGGCGCGCGCGCTGCGCGCTACATCTCCACGAGGTC	8636581
QY	828	ValThrGluLeuSerLysMetValValArgGluGluIleAspGluAsnSerThrValTyr	847
Db	8636582	GAGACACTGTCGAGCGCCCTGTCGCGCGGACGTCGAGCGGCGGACGCGGCTCCG	8636641
QY	848	IleAspAlaGlyAlaGlyAspLeuValTyrArgValGluSer-----	861
Db	8636642	GTCCAGCCCGGAGCAGGAGAGTGTGTGT-CACCTACGACCGCGGAGGACGTGAAGG	8636700
QY	862	-----GlyGlyLeu-----ValAspAlaSerThr-----GlyLys	871
Db	8636701	AGCGTGGCGGATGAGCACGATGAGCGCGACGCGTCAGTGTTCGAACTGCGGCGCGC	8636760
QY	872	LysSerAspValLeuIleHisIleAlaAsnGlyProLysArgSerAspAlaAlaGlnAla	891
Db	8636761	ACCAACCGGTCGCGCGCGGAGGCGGCCCAAGTAGTCGGCCACTGCAGACGACCC	8636820
QY	892	ValLysLysMetArgIleGluGluIleGluAspAspAsnGluGluMetIleGlu	910
Db	8636821	CTACCGTGGATG-----GTGACCGCGAGCAGCACCATCTTCACCGAGGTCGCGAG	8636871

RESULT 10

Sequence 44175, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 44175

LENGTH: 2571
TYPE: DNA

ORGANISM: Rhodospseudomonas palustris
-10-369-493-44175

Alignment Scores:

Seq. No.:	Length:	Score:
1	5.52e-177	2571
2	2264.50	458
3	71.58%	159
4	53.13%	234
5	49.77%	11
6	12	7

-09-812-350-17 (1-911) x US-10-369-493-44175 (1-2571)

1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
1 ATGAACGTTGAATAATATACCGAAGTGTGGCGGCTTCATCCAGTCAGCGCAATCGCTG 60
21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
61 GCGATGGCGGAGGCGCCATCAGCAGTCTCCGCGCTGCACATTCGAAAGTTCTGCTCGAC 120
41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGluAsnAlaAlaGln 60
121 GATTCGGAAGGCTCGCGGTGTGTCATCGACCGCGCGCGCGCAATTCGCGTGGC--- 177
61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSer---ProPro 79
178 ---ATCCTGAAGCGACCGAGAGGCGCTCGGCAGATGCCGAAGGTGTCGCGCTCGGC 234
80 ProAspPheProAlaSerSerLeuIleLysValIleArgAlaGlnAlaAla 99
235 GCGCGGCAAGCTATCTGGCCCGCGGACCGCGCGCGGCTTCGACGCTCGGAGAGGCG 294
100 GlnLysSerArgGlyAspThrHisLysAlaValAspGlnLeuIleMetGlyLeuLeu--- 118
295 GCGGAAAGCGGCGGACGCTTCGTCACCGTCGAGCGGCTCTCTCGCGCTGTGCTC 354
119 ---GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgVal 137
355 GATAAGACAGACAGCGCGCTACGCTGCTCACCAGGCGGCGGTTCACCCCGCAGAACCTC 414
138 LysSerGluValGlnLysLeuArgGlyLysGluLysValGlyLysValGluSerAlaSerGly 157
415 AATGCGGCCCATCAACGCCCTGCGC-----AAGGCGGTACCGCGGATTTCGCGAGCGGC 468
158 AspThrAsnPheGlnAlaLeuLysThrThrGlyArgAspLeuValGluGlnAla----- 175
469 GAGAACGCTATGACCGCGTGAAGAAATACGCCCGCGACCTCACCCAGCGCGCACCGGAC 528
176 GlyLysLeuAspProValIleGlyArgAspGluGluIleArgArgValValArgIleLeu 195
529 GGCAAGCTCGACCCCGTGTGCGCGCGGAGGAAATTCGCGCGACCATCCAGGTTCTG 588

QY 196 SerArgArgThrLysAsnAsnProValLeuIleGlyLysProGlyValGlyLysThrAla 215
Db 589 TCGCGCGCACCAAGAACACCGGCTGCTGATCGCGCAACCGCGGCTCGGTAAAGACCGG 648
QY 216 ValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAsp 235
Db 649 ATCGTCGAGGCTTGGCGCTGCGCATTCACCGCGACGTCGCCGAGAGCCTTGAAGGAC 708
QY 236 ValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGlu 255
Db 709 AAGAAGCTGCTGCGCTGACATGCGCGCTGATTCGCGGTGCGAAGTATCGCGCGAG 768
QY 256 PheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluLysValIle 275
Db 769 TTCAAGACGCGCTGAAGCGCTGCTCAACGAGGTTCCCGCGCGCGGCGGATCATC 828
QY 276 LeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAsp 295
Db 829 CTGTTTCATCGACGATGCACACCTGCTGGTGGCGCGGCAAGCGCGGCGGATGAC 888
QY 296 AlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThr 315
Db 889 GCGTCGATCTGCTCAAGCCCGCTGCTGCGCGAGTGCATGCTGCGCGCGGCGGAC 948
QY 316 ThrLeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGln 335
Db 949 AGCTCGATGATATCGACAGCCTCGAGAGGAGCGCGCTGCGCGCGGCGCTTCAG 1008
QY 336 GlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLys 355
Db 1009 CCGGTGTTGCTGCGCGCGGCTGCGGACACGCTCTGATCTCGCGCGGCTGAAG 1068
QY 356 GlnLysTyrGluGlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAla 375
Db 1069 GACAATACGACGACACACCGCGCTGCGCATCGCGACTCGCGGCTGCTGCGCGCGCTC 1128
QY 376 GlnLeuSerAlaAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuVal 395
Db 1129 AGCTTTTCCACCGCTACATCAGCGCGCTCTCCCGCAGAGGCGATCGACTGATG 1188
QY 396 AspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsn 415
Db 1189 GACGAGGCGCGCGCGCTGAGATGAGATGCTGACTCCAAAGCCGGAAGAACTCGACTCG 1248
QY 416 LeuGluArgLysArgMetGlnLeuGluIleGlnLeuHisAlaLeuGluArgGlyLysAsp 435
Db 1249 ATGACCGCGAGATCGTGGCTGAAGATCGAGGAGGCGCTTGAAGAGGAACCGCAT 1308
QY 436 LysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspLeuArgAspLys 455
Db 1309 CCGGCTCGAAGGCGCGCTGTCAGCTGGAGAGGAGCTTCCCGATCTCGAGAGAGAG 1368
QY 456 LeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArg 475
Db 1369 TCGCGCGGCTTGAAGCGGCTTGGAGTGGCGAGAGAACCAAGCTGTCGATGCCAGAAG 1428
QY 476 LeuLysGlnLysArgGluLeuMetPheSerLeuGlnAlaGluArgArgTyrAsp 495
Db 1429 CTGAAGAGCGAATCGACGCTCTCGGATTTGAATCTGCCAACCGCGCGCGCGGTGAG 1488
QY 496 LeuAlaArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAla 515
Db 1489 TATCAGCGCGCGCGAGCTGCTGATGCGCGGATTCGCGAGCTTGAAGAGAGATCGCT 1548
QY 516 GlnLeuGluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGlu 535
Db 1549 GAGATCGAA---GCCACGAGAACTCGCGCGGCTGATGAGGAGGCGCGCTCACCGCAAT 1605
QY 536 HistLeuAlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsn 555
Db 1606 CATATCCCGGAGTGTGCTGCGCTGACCGCGCTGCGCGGTGCAAGAGATGCTCGAAGGC 1665

А. А. БУДУНОВ, ВАРЬЯ С.

236 ValArgLeuIleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGlu 255
709 CGCAGCTAAATTTCTTAGATATGCTGTTGATTCGGGGGCAAAATTCGGGGTGA 768
256 PheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyValLe 275
769 TTTGAAGAACCCCTGAAGCAGTATTAAGAAGTTACTGAATCTGGCGCAATATTGT 828
276 LeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluLysSerMetAsp 295
829 TTAATTATTGATGAATTCACACCGTTGTTGGCGCTGCTGCAACCCCAAGCGCGATGAT 888
296 AlaAlaLeuLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThr 315
889 CGCGGTAACTTGTAAACCCGATGTTGGCGGGGTGAATTCGCTGTATTGGGGCGACA 948
316 ThrLeuGluGluThrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGln 335
949 ACTCTAGATGAATACCGCAACATATCGAAAGAGTCCGCACCTAGAAAGACGCTCCAG 1008
336 GlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLys 355
1009 CAGGTTTATCTCGATCAACCTAGTGTAGAGATAGTATTCGATTTTGGCGGGTGGAGA 1068
356 GluLysTyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaLa 375
1069 GAACGTTATGAACACCCACCGGGGTAAATTTCTGATAGTCTTTGGTCGCGCGGCC 1128
376 GlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuVal 395
1129 GTATTGTCGATCGATATATTAGCGATCGCTTCTTACTCTGATAAGCCATTGCTTGGTA 1188
396 AspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsn 415
1189 GAGCAACCCCGCCGAGATTAATAATGGAGATCACCTCCAAACAGAAAGAACTCGACGA 1248
416 LeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGlyLysAsp 435
1249 ATTGATCGCAGATTCTGCAATTTGGAATGGAGAGCTATTCGCTGCAAAAGAAAGCGAT 1308
436 LysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLys 455
1309 CGCGCTTCTCGTGAACGCTTAGAAGACTAGAAAAGAAATTCGCCGATCTCAAGAGAA 1368
456 LeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArg 475
1369 CAAGAACCCCTAAATCTCAATCGCAATCTGAAAAGATATCAATTCACAAAATTCATCC 1428
476 LeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAsp 495
1429 GTTAAAAAGAGATTGAACGGGTCAATTTAGAGATTCAGCAACAGACAGACCGATTACGAC 1488
496 LeuAlaArgAlaAlaAspLeuArgTyrGlyAlaIle-----GlnGluValGlu 511
1489 CTTAACCGACTCGGAGTTGAATACGGTAATTTAACTAGTTTGCATCTCGTCAAGTTGAA 1548
512 SerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAsnValMetLeuThrGluAsn 531
1549 GCAGTAGAGCTGAATTTGGCAAGTCCCAAGAGTGGAAAATCACTATACGGGGAAGAA 1608
532 ValGlyProGluHisIleAlaGluValValSerArgTyrThrGlyIleProValThrArg 551
1609 GTCACAGAGCTCATATTCTGAAATATTCTAATGACAGGAATTCCTCATCAGCAAG 1668
552 LeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgVal 571
1669 CTGGTGAATCTCAGAAAGAGAACTACTGCAATTTAGAGATGAACCTAGCCACCGCTGTG 1728
572 ValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGly 591
1729 ATTGGACAGAGACAGCAGTCACAGCGGTAGCGATTCGCAATTCAGCGATCGCGCGTGA 1788
592 LeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGly 611

1789 CTGGCGGATCCCAATCGTCCCATCGTAGCTTTATTTCTTGGGCGCTACGGGTGGGT 1848
612 LysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuVal 631
1849 AAAACCGAGTTGCGGAAGCGCTGGCGGCTATATGTTTCGATAGCAAGATCGCGTGGTG 1908
632 ArgIleAspMetSerGluTyrMetGluGlnHisSerValSerArgLeuIleGlyAlaPro 651
1909 CGAATCGATATGTCGGAATATATGGAGAAACACCGCGTCTCCCGTTTAAATCGGTGCGCT 1968
652 ProGlyTyrValGlyHisGluGluGlyGlnLeuThrGluAlaValArgArgPro 671
1969 CCAGGATATGTCGGTTACGAAGAGCGCAACATAACAGAGCGGATTCGCCCGCGCTCT 2028
672 TyrCysValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThrLeu 691
2029 TACTCAGTGAATCTCTTTGACGAAATCGAAGAACACACCCCTGATGTTTAAATATCTTC 2088
692 LeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArg 711
2089 TTGCAATTTCTGATGATGTCGCTGCTACTGATGCTCAAGGTCTAAGGTGAGCTTCAAG 2148
712 AsnSerValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThr 731
2149 AATGCTATTATTATCATGACTAGCAACATCGTTCGCAATACATTTCTTTGATGTCGCTGG 2208
732 GlyLysValThrMetGluValAlaArgAspCysValMetArgGluValArgLysHisPhe 751
2209 GATAATGCTACTAGCAAGAAATGCGCGTCGAGTCATGGAAGCGATCGGAATAGCTTC 2268
752 ArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspProLeuSerHisAsp 771
2269 CGTCCAGAAATCTCTCAACCGGATTGACGAATCATCATCTTCCCGTTTAGATAAGAG 2328
772 GlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAlaGlyArg 791
2329 GAATTTGCGCGAGATTGCTGCTGTCAGATCAAAAGATTACGCCAAAGATTGCGCGATCGC 2388
792 GlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAsp 811
2389 AAATATCCCTCAAGCTCTCAGATGCTGCACTTTCAGCTTTTAGCAGAGTAGGATATGAC 2448
812 ProValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValValThrGluLeu 831
2449 CCAGTTTATGGCGCGCTCCACTGAACGGGCGGATTACGCGAGAGCTAGAACTCAAAAT 2508
832 SerLysMetValValArgGluGluIleAspGluAsnSerThrValTyrIleAsp 849
2509 GCCAAAGCCATCTTGGCGGGTGAATTCACCGATGGCAACACCATCTTTGTAGAT 2562

RESULT 12

US-10-398-221-5/c
; Sequence 5, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 US
; CURRENT APPLICATION NUMBER: US/10/398, 221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 5
; LENGTH: 43980
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:

NAME/KEY: misc feature
LOCATION: (1). (end)
OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
10-398-221-5

Assignment Scores:

7.43e-174	Length:	43980
2245.50	Matches:	460
70.59%	Conservative:	193
49.73%	Mismatches:	237
49.35%	Indels:	35
12	Gaps:	12

09-812-350-17 (1-911) x US-10-398-221-5 (1-43980)

[illegible]

Db	10301	TTGCTCTTTATAGATGAAATTCATCAATTTGTCGGCGCAGGTAAACACATGACGAGCTATG	10242
QY	295	AspAlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCystileGlyAla	314
Db	10241	GATGCTGGAAACATGTTGAACCTATGCTGTGCAAGAGGTGAACCTTCATCTGTATCGCGCCA	10182
QY	315	ThrThrLeuGluGluTyrArgLysTyrValGlnLysAspAlaAlaPheGluArgArgPhe	334
Db	10181	ACAACCTTGGACCAATACCGCCATATATCGAAAAAGATGCTGCTACATAGAAGACGTTTC	10122
QY	335	GlnGlnValTyrValAlaGlnProSerValProAspThrLieserLieserLieserLysLeu	354
Db	10121	CAAAAGTAGTCTGTTCCAGAACCAACCGGTGGAAGATACAGTTTCCATTTTACGTGTTTA	10062
QY	355	LysGluLysTyrGluGlyHisGlyValArgLileGlnAspArgAlaLeuLeuAsnAla	374
Db	10061	AAAGAACGTTTGAATCCATCATGGGTAAATATTCATGATAATGCTTATGTTGCGCCA	10002
QY	375	AlaGlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeu	394
Db	10001	GCTAGCGCTTCTTAATCGTTTACATACCGATCGTTTTTTTACCGGATAAAGCAATTTGATT	9942
QY	395	ValAspGluAlaCysAlaasnValArgValGlnLeuAspSerGlnProGluGluIleAsp	414
Db	9941	GTAGATGAAGCGGTGCGCACTTATTCGTGCAAAATTGACTCTATGCCAAGTGAATTCAT	9882
QY	415	AsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLys	434
Db	9881	GAAGTAACAAGAAAAAGTCATGCAGCTAGAAATTTGAGNAGCGGCATTTAAAGAAAGAAAA	9822
QY	435	AspLysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAsp	454
Db	9821	GACCCAGCAAGTGAACGGCGCTTAGAGATGTTACAGCGAGAACTAGCTGACTATAAAGAA	9762
QY	455	LysLeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArg	474
Db	9761	GAAGCAAAATAAATGAAGTCTAAATGGGAATCCGAAAAAGCCGAAATCAGTAAAAATCGT	9702
QY	475	ArgLeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyr	494
Db	9701	GAAGTTCGTGAACAAATCGATCATCTGGGTATCGAATTTAGAGACGACGAGAACCACTAC	9642
QY	495	AspLeuAlaArgAlaAspLeuArgTyrGlyAlaIleGlnGluValGlnSerAlaIle	514
Db	9641	GATTTAAATATAAGCTGCTGAACTTCGCCACGTTAAATTCAGCAGTAGAAAAAGAAATTA	9582
QY	515	AlaGlnLeuGluGlyThrSerSerGlu-----GluAsnValMetLeuThrGlu	530
Db	9581	TTAGCATTTAGAAAACGGAAATTCGCGAAAAAACACGCCCAAGATCGAATTTTACAAAGAA	9522
QY	531	AsnValGlyProGluHisIleAlaGluValValSerArgTyrThrGlyIleProValThr	550
Db	9521	GAAGTAACAGAAAAATGAATTCGCTGAATTTGTGGACATGGACTGGAATACCGAGTACT	9462
QY	551	ArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyAlaAspArgLeuHisLysArg	570
Db	9461	AAGTTAGTAGAAGGAGAACGCGAAAAACTGCTTAATTAGCCGATGTTCTTCATCAAAAA	9402
QY	571	ValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAla	590
Db	9401	GTAATTTGTCGAAGACGCGGTTCAGTTAGTTAGTGTGATGTCAGTATTTAGCTGCTGCTG	9342
QY	591	GlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyVal	610
Db	9341	GGGATTAAGATCCAAAACGACCAATCGGATCTTTATTTTCTTAGGCCCAACCGGTGTT	9282
QY	611	GlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeu	630
Db	9281	GCTAAAAACGGAATCAGCAAGCATTTAGCCTATAATATGTTTGAATCTTGAAGATCATATG	9222
QY	631	ValArgIleAspMetSerGluTyrMetGluGlnHisSerValSerArgLeuIleGlyAla	650
Db	9221	ATTGCAATGATATGTCGAATACATCGAGAACATTCGCTATCAAGATCTTCGGGCTG	9162

```

651 ProProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArg 670
9161 CTTCCAGGTGTAITGGATATGAGGAGCGGACCACTTACGAGCTGTGAAGCAAC 9102
671 ProTyrCysValLeuLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThr 690
9101 CCATATTCGATGTCCTTACTTATGATCAATCGAAATCGAAAGCCCATCCAGACGCTATTTAATATC 9042
691 LeuLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPhe 710
9041 CTTTACAGTACTTGACATGTCGGATACGATTCACAAAGGCGCTTAATGACTTT 8982
711 ArgAsnSerValLeuLeuMetSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeu 730
8981 AAAAACACCGTAATCATATGACCTTAATATCGCTCTAATTTATTACITGAAGAACC 8922
731 ---ThrGlyLysValThrMetGluValAlaArgAspCysValMetArgGluValArgLys 749
8921 GAAGAAGGTGAATCTCGCTGAACTAGATCAGAC---GTAATGCAAAATTTTGCRAAGT 8965
750 HisPheArgProGluLeuLeuAsnArgLeuAspGluLeuValValPheAspProLeuSer 769
8864 GAATTTAAACCAAGATCTTAAACCGGTAGATGATATTAATTCATTTAAACCACTTACA 8905
770 HisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAla 789
8804 CTCGCTGATATTAAAGGCATGTGCAAAATATTAGTAGAAGCACTTCAATCGCCTGCT 8745
790 GluArgGlyValAlaLeuAlaValThrAspAlaLeuAspTyrLeuLeuAlaGluSer 809
8744 GATCAGAAATACCATTAATTTTACAGTAACGCAAAAGCCCTTATTGCAAGAAGCT 8685
810 TyrAspProValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysValValThr 829
8684 TATGACCAAGTATACGCTGCGCTCCACTGAACGATATATTGTGCGTATGCGAAGC 8625
830 GluLeuSerLysMetValValArgGluGluLeuLeuAspGluAsnSerThrValTyrIleAsp 849
8624 CTTCTTCAAGAAATTTGATCTGTGTAAATATTGCGCATCTCTCCGTTGAATTTGAT 8565
850 AlaGlyAlaGlyAspLeuValTyrArgValGluSerGlyLeuValAspAlaSerThr 869
8564 TTACAAGATAAGAAATTTACTTTTAAAGT-----ACA 8532
870 GlyLysLysSerAspValLeuLeuHisIleAlaAsnGlyPro----- 883
8531 GAATAACATGAAATGCTTTGATTTCAGCTCGAAGGGAACCTTTTGTGATTAGTC 8472
884 -----LysArgSerAspAlaAlaGlnAlaValLysLysMetArgLeuGluLeuGlu 901
8471 GTTTAGAAATAAGCGAAATATGGAAGG---ATGAGGAAGATGAAATTAGTATTATTCGT 8415
902 AspAspAspAsnGlu 906
8414 CATGGTCAAAGTCAA 8400

```

SULT 13

-10-398-221-2058/c

Sequence 2058, Application US/10398221

Publication No. US20040018514A1

GENERAL INFORMATION:

APPLICANT: KUNST, Frederik

APPLICANT: GLASER, Philippe

TITLE OF INVENTION: Listeria innocua, genome and applications

FILE REFERENCE: 344 702 - US

CURRENT APPLICATION NUMBER: US/10/398,221

PRIOR FILING DATE: 2003-03-27

PRIOR APPLICATION NUMBER: PCT/FR 01/03 061

PRIOR FILING DATE: 2001-10-04

PRIOR APPLICATION NUMBER: FR 00/12 697

PRIOR FILING DATE: 2000-10-04

NUMBER OF SEQ ID NOS: 4025

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2058

; LENGTH: 3011208

; TYPE: DNA

; ORGANISM: Listeria innocua

US-10-398-221-2058

Alignment Scores:

```

Pred. No.: 1,23e-171 Length: 3011208
Score: 2245.50 Matches: 460
Percent Similarity: 70.59% Conservative: 193
Best Local Similarity: 49.73% Mismatches: 237
Query Match: 49.35% Indels: 35
DB: 12 Gaps: 12

```

US-09-812-350-17 (1-911) x US-10-398-221-2058 (1-3011208)

```

QY 1 MetAsnProGluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20
Db 2343102 ATGGATTACAAAATTTTACCACACAGTCCAGCAACAAATTCGCGACGCAAAATCTA 2343043
QY 21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSer 40
Db 2343042 GCTATTGCATCAGAACATCAAGAAATTCAGCGTTATACATGTTTAAAGTGTATTAACT 2342983
QY 41 AspProThrGlyLysPheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60
Db 2342982 GAG---AGTGACTTTGCAAAACGCGCATATGATGTAGCA-----GAAGTTAATGTTGAG 2342932
QY 61 SerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuPro-----SerGlnSerPro 78
Db 2342931 GCACCTTCAAAAGTAGTAGACGAATCTTTAAGAAAAATCTCTGTTGTTTCAGGAAGCGGA 2342872
QY 79 ProProAspAspIleProAlaSerSerSerLysLysValIleArgArgAlaGlnAla 98
Db 2342871 GTGAATATGCGCAAGCAATGAGTCAAGCACTTTTCAACATAATGAGAGATGCGAGAAAA 2342812
QY 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118
Db 2342811 GAACAAAAACAATTAGACGATGATTTGTATCGACAGAACATCTTATCTTAGCTGTTAATG 2342752
QY 119 Glu-----AspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArg 136
Db 2342751 GACCAAAAGTCAAAACCCCATCACAGTCGAATTTAAAGCAAAATAATCTTAAAAACAA 2342692
QY 137 ValLysSerGluValGluLysLeuArgGlyLysGluGlyLysLysValGluSerAlaSer 156
Db 2342691 ATTACGAAGCAATTTTAAATAATCAGAGGA-----GGGAAAAAAGTGACTTCTCAAAAT 2342638
QY 157 GlyAspThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGln----- 174
Db 2342637 GCAGAGGAAACTATGAGCTTTAACAAATATGACGAGATTTAGTCGCGGAAGTAAGA 2342578
QY 175 AlaGlyLysLeuAspProValIleGlyArgAspGluGluIleArgArgValValArgIle 194
Db 2342577 AGCGGAAACTTGATCCAGTGATTCAGCGACGACGAGAAATTCGTAATGTCATCCGAAT 2342518
QY 195 LeuSerArgArgThrLysAsnAsnProValLeuIleGlyProGlyValGlyLysThr 214
Db 2342517 TTATCAAGAAAAACAAAAATAATCCAGTTCTAATTTGGTGAACCAAGCTGTTGTTAAACA 2342458
QY 215 AlaValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThr 234
Db 2342457 GCATTTGAGAGGTTTAGACACACGTTTGTGCAAGAGACGTTCCAGAGGATTTGAAA 2342398
QY 235 AspValArgLeuLeuSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGly 254
Db 2342397 GATAAAACAATTTATTTCCCTTGATATTGTTGTTCCCTTATTGCTGGGCTAAGTATCTGCT 2342338
QY 255 GluPheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysVal 274
Db 2342337 GAATTTGAGAACGTTTAAAGCAGTACTCCAAAGAGTAAACAAAGCGATGACAAAT 2342278

```

275 IleuPheIleAspGluIleHisLeuValLeuGluYalaGlyLysThrGluSerMet 294
2342277 TTGCTCTTTATAGATGAATTCATACAAATGTGCGGCGAGGTAAACAGATGAGCTATG 2342218
295 AspAlaAlaAsnLeuPheLysProMetLeuAlaArgGlyClnLeuArgCysIleGlyAla 314
2342217 GATGCTGGAAACATGTTGAACCTATGCTTGCAGAGGTGAATTCACCTGATTCGGCGCA 2342158
315 ThrThrIleuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPhe 334
2342157 ACAACTTTGACCAATACCGCAATATATCGAAAGAGATGCTGCACAGAAAGACGTTTC 2342098
335 GlnGlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeu 354
2342097 CAATAAGTACTGTTCCAGAACCAACCGTGAAGATACAGATTTCCATTTTACGTGCTTTA 2342038
355 LysGluLysTyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAla 374
2342037 AAAGAACGTTTGAATCATCATGGGTAAATATTCATGATTAATGCCTTAGTCCCGCA 2341978
375 AlaGlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeu 394
2341977 GCTAGCTCTTCTAATCGTTACATACGGATCGTTTATACGGATAAAGCAATTCATTTA 2341918
395 ValAspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAsp 414
2341917 GTAGATGAAGCGGTGCAACTATTCGTGCGAAATGATCTATCCCAAGTGAATTCAT 2341858
415 AsnLeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGluLys 434
2341857 GAAGTAAACAAGAAAGATCATGCAGCTAGAAATTTGAAGCGCGCATTTAAAGAGAAAGAA 2341798
435 AspLysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAsp 454
2341797 GACCAGCAAGTGAAGCGGCTTACAGATGTTACAGCAGAACTAGCTGCTACTATAAGAA 2341738
455 LysLeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgLysGluIleArg 474
2341737 GAAGCAATAAATGAAGTCTTAATGGGAATCCGAAAGAACGAAATCAGTAAATTCGT 2341678
475 ArgLeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyr 494
2341677 GAAGTTCGTGAACAAATCGATCATCTCGTTCATGATTTAGAAAGCAGAGAAACAATAC 2341618
495 AspLeuAlaArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIle 514
2341617 GATTAAATAAAGCTGCTGAATTCGACCGTGAATTTCCAGCAGTAGAAGAAAGAAATTA 2341558
515 AlaGlnLeuGluGlyThrSerSerGlu-----GluAsnValMetLeuThrGlu 530
2341557 TTAGCATTAGAAACGAAATTCGCGAAAGAAACACGCCCAAGAGATCGAATTTTACAAGAA 2341498
531 AsnValGlyProGluHisIleAlaGluValValSerArgTyrThrGlyIleProValThr 550
2341497 GAAGTAAACAGAAATGAATCGCTGAAATTTGTCAGACGTGACTGGAATCCAGTGACT 2341438
551 ArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArg 570
2341437 AAGTTAGTAGAAGAGAACCGGAAATCTGCTAAATAGCCGATGTTCTTCATCAAAA 2341378
571 ValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSerArgAla 590
2341377 GTAATTCGTCAAGACGCGGTTCAGTTAGTAGTACGATGATGATTCGTGCTGCG 2341318
591 GlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGlyProThrGlyVal 610
2341317 GGGATTAAAGATCCAAACACCAATCGATCTCTTTATTTCTTAGGCCCAACCGGTGTT 2341258
611 GlyLysThrGluLeuAlaLysAlaLeuAlaGluClnLeuPheAspAspGluAsnLeu 630
2341257 GGTAAACGGAATAGCAAGCAATAGCCATAATATGTTGATTCAGAGATCATATG 2341198
631 ValArgIleAspMetSerGluTyrMetGluGlnHisSerValSerArgLeuIleGlyAla 650

Db 2341197 ATTGCAATGTATGTCTGAATATACATGAGAAACATTCGCTATCAAGACTTGTGCGGCT 2341138
QY 651 ProProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArgArgArg 670
2341137 CCTCAGGTTATGTTGATATAGAGAGCGGCAACACTTACCGAGCTGTAAAGACGCAAC 2341078
QY 671 ProTyrCysValIleLeuPheAspGluValGluLysAlaHisValAlaValPheAsnThr 690
2341077 CCAATTCGATGTTGCTTACTTGTGATGAATCGAAAGCCATCCAGACGCTATTATATATC 2341018
Db 691 LeuLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyArgThrValAspPhe 710
2341017 CTTTACAGTACTTTCAGCATGTCGATACGATTCACAGAGCGCTTAATATGACTTT 2340958
QY 711 ArgAsnSerValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeu 730
2340957 AAAAACACCGCTAATCATATGACGCTAATATCGCTCTAATTTACTTTGAAGAAC 2340898
Db 731 ---ThrGlyLysValThrMetGluValAlaAlaArgAspCysValMetArgGluValArgLys 749
2340897 GAAGAGGTGAATCTCGCTCGAATAGAAATCAGAC---GTAATGCAAAATTTGCAAGT 2340841
QY 750 HisPheArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspProLeuSer 769
2340840 GAATTTAAACACGAAATCTTAAACCGGTAGATGATATTAATCTATTTAAACCACTTACA 2340781
Db 770 HisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeuAla 789
2340780 CTCGCTGATATTAAGCAATTTGCGAAATTTAGTAGAAGAACTTCAAATTCGCTTGT 2340721
QY 790 GluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGluSer 809
2340720 GATCAGAAATTCAGATTAATCTATTCAGATAACGCAAGCCCTTTATTCAGAGAAAGCT 2340661
Db 810 TyrAspProValTyrGlyAlaArgProIleArgTyrMetGluLysLysValValThr 829
2340660 TATGACCCAGTATACGTCGCGCTCCCTCAGTAAAGATTTATGCGCATTTTCCGTTGAT 2340541
QY 850 AlaGlyAlaGlyAspLeuValTyrArgValGluSerGlyGlyLeuValAspAlaSerThr 869
2340540 TTACAAGATAAGAAATTTACTTTTAAGTC-----ACA 2340508
Db 870 GlyLysLysSerAspValLeuIleHisIleAlaAsnGlyPro-----883
2340507 GAATAACATGGAATGTTTGTATTCAGCTCGAAGGGAAGAACTTTTGTGATAGTC 2340448
QY 884 -----LysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluLeu 901
2340447 GTTTGAAATTAAGCGAAATGGAAG---ATGAGGAAGATGAAATAGTATTATTAATCGT 2340391
QY 902 AspAspAsnGlu 906
Db 2340390 CATGTCGAAGTGAA 2340376

RESULT 14

US-10-369-493-46242
; Sequence 46242, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10 (52052) B
; CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 46242
 LENGTH: 2715
 TYPE: DNA
 ORGANISM: Schizosaccharomyces pombe
 S-10-369-493-46242

Alignment Scores:

red. No.: 2,06e-174 Length: 2715
 core: 2234.50 Matches: 460
 Percent Similarity: 66.52% Conservatives: 158
 Percent Local Similarity: 49.52% Mismatches: 258
 Indels: 53
 Gaps: 11

S-09-812-350-17 (1-911) x US-10-369-493-46242 (1-2715)

6 PheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeuAlaValAsnAlaGly 25
 16 TTTATCTGCAAAAGCGCAAGACATTGCTGATGCTGCTCAATTCGACAAATCTTATGGT 75
 26 HisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuIleSerAspPro-----42
 76 CATTACAAATTAACCCCTATTACATTGCTGCTCTTTTGTTCGACAGTCAGTAAAC 135
 43 ---ThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGlnSer 61
 136 GGTACTACTACTGCGCACCATTCTGCACAAAGCTGGTGGTAT-----GGTCAAAAG 189
 62 AlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerProProAsp 81
 190 TTTGAGCGAAGTGTACTAGTCGGTGTAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 249
 82 AspileProAlaSerSerSerLeuIleLysValIleArgAlaGlnAlaAlaGlnLys 101
 250 CAAGTGAATCTGTCACCTGAAAGTGGCAAGCTGCTGCGAATGCTCATGAACTTCAAAA 309
 102 SerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeuGluAspSer 121
 310 ACACAAAAGATTTTATGCGAAGCTCAAGATCATTTATCGCTGCTGCTGCTGCTGCTGCT 369
 122 GlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLysSerGluVal 141
 370 ACGCTAAAGATTTATGGCAGAAGCGGTGTCACCCGAAAGCGTTTGAATTTGCTGTA 429
 142 GluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAspThrAsnPhe 161
 430 AATAACGTCGCTGCT-----AATAAAGAATCGATTCAAGATCCGAGGAGGATTT 483
 162 GlnAlaLeuLysThrThrGlyArgAspLeuValGluGlnAla-----GlyLysLeuAsp 179
 484 GATGCACTAAATAAATTTACTGTTGATCTTACAGAGCTGCTGATAAATGCTCAACTGGAT 543
 180 ProValIleGlyArgAspGluGluIleArgValValArgIleLeuSerArgGlyThr 199
 544 CCAGTTATCGACGTGAGATGAATTCGTCGACAAATTCGCTGCTGCTGCTGCTGCTGCTGCT 603
 200 LysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGly 219
 604 AAAAATAATCCAGTTCTTATTTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 663
 220 LeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArgLeuIle 239
 564 CTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 723
 240 SerLeuAspMetGlyAlaLeuValAlaGlyValLysThrArgGlyPheGluGluArg 259
 724 AGCCTGAGCTGCGAAGCTTAGTTGCGGCTCCAAATTTGCTGGTGAATTTGAAGAGCGT 783
 260 LeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPheIleAsp 279

Db ATCAAACTGTTTAAAGAAAGTTGAAGAAAGTAAACTCCCATTTATTATTCGTAGAT 843
 QY GlnIleHisLeuValLeuGlyAlaGlyLys---ThrGluGlySerMetAspAlaAlaAsn 298
 Db GAAATGATCTTTTAAATGCGCGAGGTTCCGGTGGTGAAGTGGCATGGACGCTGCCAAT 903
 QY LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGlu 318
 Db TTGCTGAAGCAATGCTTGCACGTTGTAAGCTTCACTGTATTGGTGTACTACTTTGGCT 963
 QY GluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgPheGlnGlnValTyr 338
 Db GAATACAAAGAAATACATCGAAAGATGCTGCTTTCGACGTAGATTCACAAATCATCTTG 1023
 QY ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyr 358
 Db GTAAAGAGCTTTCTATTGAAGATACGATTTCTATCTCTGCTGCTTAAAGAAAGTAT 1083
 QY GluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSer 378
 Db GAGTTTCATCATGCTGCTTACTATCTGACCGTGCCTTGGTTACTGCTGCCATCTCGCT 1143
 QY AlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAla 398
 Db TCCCGGTACTCACTTCTGCTGCTTCCAGATTTCTGCCATCTGCTGCTGCTGCTGCTGCT 1203
 QY CysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArg 418
 Db CGCGCTGCGAGTTCGTTTACTCTGCTGAATCTCAGCCGAAAGTTCTTGATAACCTTGAGA 1263
 QY LysArgMetGlnLeuGluIleGluHisAlaLeuGluArgGluLysAspLysAlaSer 438
 Db AAGCTTCGAACTCCGCGTGTAAATTCGTGCTTTAGACGCTGGAAGAGCAAGCTTCC 1323
 QY LysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnPro 458
 Db AAAGAAGCTCTCAAGCTGCTCTAAAGAGCTGAACAAGTGAAGAGAAAGAACAGACT 1383
 QY LeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArgLysGln 478
 Db ATTCGTGAGAGTATGAATTAAGAAAGTCTCTGCTGAGTGAAGTCAAGATGCTAAGCT 1443
 QY LysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArg 498
 Db CGTCTTGATGAGTTGAAGCTAAGCTGAGGATCCGCGCTGCTGATGATTTCACTCTT 1503
 QY AlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGlu 518
 Db GCTGCTGATTTGAATTAATCTATGTTATTCCTGATTTGCAAAAACGATTTGAATATTAG 1563
 QY GlyThrSerSerGluGluAsnVal-----526
 Db CAACAAAAAGAGCTGATGAGAGCGGATTCGAAACGCTCAACGCTCAGACCT 1623
 QY MetLeuThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThrGly 546
 Db TTATTTATGAGTAGTGGCGCGATCAATCAACCAATCGTTGCAAGTGGACTGGT 1683
 QY IleProValThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArg 566
 Db ATTCCGCTTACTCTGCTTTGAAGACCAACTGAAAGAGCAAGCTCTTAACATGGAGAGGTT 1743
 QY LeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeu 586
 Db CTGAGTAAACAGTAATGCTCAAAATGAAGAGGATCTGCTGAGTTCGCAACCGCCATCGC 1803
 QY ArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeuGly 606
 Db TTATCTGCTGCTGCTCTTTCTGATCTTAATCAACCAATCGCATCTTTCTTTGTTTCGGT 1863
 QY ProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAsp 626

```

1864 CTTTCGGTACTGCTAAGACTTTATTGACAAAGCATTGGCATCATTTATGTTGATGAT 1923
627 GluAsnLeuValArgIleAspMetSerGluThrMetGluGlnHisSerValSerArg 646
1924 GAAATGCAATGATTCGTATTGATGATGATGATGATGATGATGATGATGATGATGAT 1983
647 LeuIleGlyAlaProGlyValGlyValGlyValGlyValGlyValGlyValGlyVal 666
1984 CTATTGGAGCACCACCGGTTATGTTGGACATGAAGCTGTTGGACAGCTTACTGAACA 2043
667 ValArgArgProGlyValGlyValGlyValGlyValGlyValGlyValGlyValGly 686
2044 CTTTCGGTACTGCTAAGACTTTATTGACAAAGCATTGGCATCATTTATGTTGATGAT 2103
687 ValPheAsnThrLeuLeuGlnValLeuAspGlyArgLeuThrAspGlyGlnGlyArg 706
2104 GTATTGACTGATCTTTTGGAGTACTAGATGATGATGATGATGATGATGATGATGAT 2163
707 ThrValAspPheArgAsnSerValIleMetThrSerAsnLeuGlyAlaGluHisLeu 726
2164 GTTGTGATGCCAAGATGCTGTTATCATTTATGATCTTAACCTTGGGCGCTGAATCTTG 2223
727 LeuAla-----GlyLeuThrGlyLysValThrMetGluValAlaAlaArgAspVal 743
2224 ACAACAGACAATGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
744 MetArgGluValArgLysHisPheArgProGluLeuLeuAsnArgLeuAspGluIleVal 763
2281 ATGAATCCGATCGTGGCTTCCTTCGTCGCAATTTTGAATCGTATCTCTCAATCGTT 2340
764 ValPheAspProLeuSerHisAspGlnLeuArgLysValAlaAlaArgLeuGlnMetLysAsp 783
2341 ATCTTTAATCGTCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
784 ValAlaValArgLeuAla-----GluArgGlyValAlaAlaLeuAlaValThrAspAlaAla 801
2401 GTTCAGAAACCGCTTCAATCCCAATCATCGCTCAATTAATAATCGAGTTAGCGCAAGCT 2460
802 LeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgProIleArgArg 821
2461 AAGACCTTCCTTGGAGAGCTGTTTATTCCTGCTTATGCTGCTGCTGCTGCTGCTGCT 2520
822 TrpMetGluLysValValThrGluLeuSerLysMetValValArgGluGluIleAsp 841
2521 GTCAATCAAAATCAAGTTTGAATCCCAATGCTGCTTAAATTTGAATGGCCAGCTTGGC 2580
842 GluAsnSerThrValTyrIleAspAlaGlyAlaGlyAspLeuValTyrArgValGluSer 861
2581 GATAAGGAGACTGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2637
862 GlyGlyLeuValAspAlaSerThrGlyLysLysSerAspValLeuIleHisIleAlaAsn 881
2638 -----CATGAAGCTAAT 2649
882 GlyProLysArgSerAspAlaAlaGlnAlaValLysLysMetArgIleGluGluIleGlu 901
2650 GCAAAACGCTCTGCTGAT-----ATTGACATGGATGGATTATGAC 2688
902 AspAspAspAsnGluGluMetIleGlu 910
2689 GACGATGTTATGATGAAGATTGAA 2715

```

10-369-493-44640
 sequence 44640, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 44640
 ; LENGTH: 2588
 ; TYPE: DNA
 ; ORGANISM: SYNECHOCOCCUS SP. WH 8102
 ; US-10-369-493-44640

Alignment Scores:
 Pred. No.: 6,56e-174 Length: 2588
 Score: 2228.00 Matches: 455
 Percent Similarity: 71.88% Conservative: 166
 Best Local Similarity: 52.56% Mismatches: 221
 Query Match: 48.97% Indels: 23
 DB: 12 Gaps: 10

US-09-812-350-17 (1-911) x US-10-369-493-44640 (1-2588)

```

QY 1 MetAsnPro-----GluLysPheThrHisLysThrAsnGluThrIleAlaThrAlaHis 18
Db 1 ATGCAACCCACAGCCGACAGTTTCAACGACAGCCCTGGCGCCCATCTGTCGGCGCCAG 60
QY 19 GluLeuAlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeu 38
Db 61 CAACTGCCCGGCTCTCAGGCACAGCAGCTGGAACACCAATCTGCTGCTGCTGCTGCTG 120
QY 39 IleSerAspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAla 58
Db 121 CTTCCAGAG--AACGGACTGGCGCGGAATCCTCAGCAAAACCGCGCTTGAC----- 171
QY 59 AlaGlnSerAlaGluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78
Db 172 GTCACCACTTCGAGGCGCAGTGTGAAGCCCATCTCAACGGCTTCCAGCTTGGCTCG 231
QY 79 ProProAspAspIleProAlaSerSerSerLeuIleValIleArgArgAlaGlnAla 98
Db 232 GCACCCGACTCGGTGTTTCTTGGCCGATCCCTCAACAGCCCTCGACGCGCGAGCAG 291
QY 99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuMetGlyLeuLeu 118
Db 292 CGCGGGATGATTCGCGCAGCTTCATCGCATCGAGCATCTGCTTCTGCGCTGGCG 351
QY 119 GluAspSerGlnIle---ArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgVal 137
Db 352 GAGGATGATCGTGGCGCGCAACTGCTCAGCCAGCCGCGGGTGACCAACACACACTC 411
QY 138 LysSerGluValGluLysLeuArgGlyLysGluGlyLysValGluValGluSerAlaSerGly 157
Db 412 AAGGAGCCCATCACGCGCGTGGCGCAACCAA-----ACGGTGACAGACAGAACCCCT 465
QY 158 AspThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAla----- 175
Db 466 GAGGCCACTCAGATCCCTGGCGAATAACGCGCGCATCTCAGACAGCGCGCGCGCAT 525
QY 176 GlyLysLeuAspProValIleGlyArgAspGluGluIleArgValValArgValLeu 195
Db 526 GGTCACTGATCGCTGATCGGACGGATGATGATGATGATGATGATGATGATGATGATGAT 585
QY 196 SerArgArgThrLysAsnAsnProValLeuIleGlyGluProGlyValGlyLysThrAla 215
Db 586 AGCCCGCGCACCAAGAACCAACCCGCTGCTGATCGGTGAACCCGCGGTGCGGCAACAGCA 645
QY 216 ValValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAsp 235
Db 646 ATCTGTCGAGGGCTTGGCACAGCGGATCTCAATGATGATGATGATGATGATGATGATGAT 705
QY 236 ValArgLeuIleSerLeuAspMetGlyValAlaLeuValAlaGlyAlaLysTyrArgGlyGlu 255

```


706 CGACAGCTCATTCACCTCGACATGGTGGTCCCTGATCGCCGGGGCGAAATACCGGGGTGAG 765
256 PheGluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIle 275
766 TTCGAAGAACGGCTCAAGCGCTGCTGAAGAGGTCCACCCTCGGATGGACAGATTGTG 825
276 LeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAsp 295
826 CTGTTTCATCGATGATCCACACCGGTGTGTGGCGCTGGCGCCAGCGGTGTGCTCATGGAC 885
296 AlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThr 315
886 GCCAGCAATCTGCTGAACACCGATGCTGGCCCGGGTGAATGCGCTGCATCGGGGCCACC 945
316 ThrLeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGln 335
946 ACCCTTGATGACACCGCTCAGCATCATCGAAAGGATCCCGCCCTGGAGCGACGCTTCAG 1005
336 GlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLys 355
1006 CAGGTGTGGTGGATCAGCCACAGTGCAGGCTACGATTCGCCGACAGCGATTTCAATCTCGCGGGCCTGAAG 1065
356 GluLysTyrGluGlyHisHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAla 375
1066 GAGGCTACCGAGGTACACCGCGGTACGATTCGCCGACAGCGCTTTGGTGGCCCGGCC 1125
376 GlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuVal 395
1126 ATGCTCAGCAGCAGATCATACCGATCGCTTCTCGCCGACAGCGCATTCGACCTGGTG 1185
396 AspGluAlaCysAlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsn 415
1186 GATGAATCCGCGCCCGCTGAAATCGAATCGAATCACCTCCAAACCGGAGCAGATCGACGAG 1245
416 LeuGluArgLysArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGlyLysAsp 435
1246 ATCGATCGCAAAATCTCGAGCTGGAGATGGAAACTCTCCCTCGCCCGTGGTCCGAC 1305
436 LysAlaSerLysAlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgLys 455
1306 AGTGCCAGCAGGAGCGGTGGCAACGCGATGAACGGGATTTGGCGGAATCTGGGTGAACAA 1365
456 LeuGlnProLeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgArg 475
1366 CAGAGCAGCTCAATGCCCGAGTGGCAGAGCGAAAGGTGCCATTGATCAGTCTCGGCT 1425
476 LeuLysGlnLysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrArg 495
1426 CTCAGAGGAGATCGAACCGGTGCAGTGCAGTGGAGCAGCGCCAAACGCAATTACGAC 1485
496 LeuAlaArgAlaAlaAspLeuArgTyrGlyAla-----IleGln 508
1486 CTCAAAGGACGCTCAATCGAATACGAACTCTCGCCACGCTGCAACGCCAGCTGCAG 1545
509 GluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAsnValMetLeu 528
1546 GAGCAGAGACCTCTTGGAGACGAGCGGAACA-----GACAAAC-CTGCTG 1595
529 ThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThrGlyIlePro 548
1596 CGGAGGAGGTCCAGAGCAGACATCGTGAATGATTCGCAAGTGGACCGCGATCCCC 1655
549 ValThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAspArgLeuHis 568
1656 GTGCTCGCCTGTGAGAGCGGAAATGGAGAACTCTGCTGAGTGGAGGACGACCTGCAT 1715
569 LysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIleLeuArgSer 588
1716 CAACGGGTGATCGACAGAACCAAGCGGTCATGCGGTGGCGGCGGATTCACGCTCA 1775
589 ArgAlaGlyLeuGlyArgAlaGlnProThrGlySerPheLeuPheLeuGlyProThr 608
1776 AGGGCAGGCTCAGCGATCCCAACCGGCCCTCAGCGGCTTCTGTTCTCGGGCCGACG 1835

QY 609 GlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAspAspGluAsn 628
Db 1836 GGTGTGGCAGACGAACTGTCCAGGCGCTGGCCAAATCGCTGTTCGACAGCGACGAC 1895
QY 629 LeuLeuValArgIleAspMetSerGluTyrMetGluGlnHisSerValSerArgLeuIle 648
Db 1896 GCCATGTGTCGATCGACATGTGCGAGTACATGGAGAAACACACCGTGGAGCGCATGATC 1955
QY 649 GlyAlaProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGluAlaValArg 668
Db 1956 GGTGCCCTTCGCGGTATGTGGGTATGAAGCGCGCGCCAGTCCAGGAGCGGTACGG 2015
QY 669 ArgArgProTyrCysValIleLeuPheAspGluValGluLysAlaHisValAlaValPhe 688
Db 2016 CAGCGGCTTACCGCGTGTGATCTGTTCAGCGAGGTGGAGAAAGCCACCCCGATGTGTC 2075
QY 689 AsnThrLeuLeuGlnValLeuAspAspGlyArgLeuThrAspGlyGlnGlyVal 708
Db 2076 AATGTGATGCTCGATCTCGATGACGCGCGGTTCACGACGGCCAGCGCGCGCGGTG 2135
QY 709 AspPheArgAsnSerValIleIleMetThrSerAsnLeuGlyAlaGluHisLeuLeuAla 728
Db 2136 GATTTCACCAACACCGTGTGATCTCACCAGCAACATCGCGCAGCAATCGATTCTTGAG 2195
QY 729 GlyLeuThrGlyLysValThrMetGluValAlaArgAspCysValMetArgGluValArg 748
Db 2196 ---CTGGTGGGATCCGAGCAACACACGCGGATGGAGCAACGGGTCAACGAGCGGCTC 2252
QY 749 Lys---HisPheArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspPro 767
Db 2253 AAGGCCAAATTCGTCGCCGAATTTCTCAACAGGCTCGACGATCAGATCATCTTCGCGACG 2312
QY 768 LeuSerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArg 787
Db 2313 CTTGAAAAGAGGAGCTCGCGCGATCGTGTCACTGCAGGTGGAGCGTCTGCGGTGCGCA 2372
QY 788 LeuAlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAla 807
Db 2373 CTGGAGCAACCGAAGCTGGACCTTCAGCTCAGCGCAATCGCGCGGATTCGTCGGCCACC 2432
QY 808 GluSerTyrAspProValTyrGlyAlaArgProIleArgArgTyrMetGluLysLysVal 827
Db 2433 ATTGCTTTGACCGGTTTACCGAGCAAGACCCCTCAACCGCGCATCCAGCGGAGCTG 2492
QY 828 ValThrGluLeuSerLysMetValValArgGluGluIleAspGluAsnSerThrValTyr 847
Db 2493 GAACACCAATCGCCAAAGCGATCTCTGGCGGACAGCTCAGTGGAGGACAGCGGTGCG 2552
QY 848 IleAspAlaGly 851
Db 2553 GTTGATGCGAGT 2564

Search completed: February 13, 2004, 10:47:34
Job time : 14451 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - nucleic search, using frame_plus_p2n model

on: February 13, 2004, 01:40:31 ; Search time 3827 Seconds
(without alignments)
5785.570 Million cell updates/sec

le: US-09-812-350-17

fect score: 4550

quence: 1 MNPEKTKHTNETHATAHEL.....VKOMRIEDDDNEEMIED 911

oring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

urched: 22781392 seqs, 12152238056 residues

al number of hits satisfying chosen parameters: 45562784

imum DB seq length: 0

imum DB seq length: 2000000000

t-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

mand line parameters:

DEL=frame+ p2n.model -DEV=xlp
/cgm2_1/USPTO_spool_p/US09812350/runat_10022004_101515_13361/app_query.fasta_1.1095
=EST -QFMT=fastap -SURF=p2n.rst -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
ITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
CALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
TPWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
ER=US09812350 @CGN_1_1.2874 @runat_10022004_101515_13361 -NCPU=6 -ICPU=3
MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
V TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
APEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

abase :

EST:*

- 1: em estba:*
- 2: em esthum:*
- 3: em estin:*
- 4: em estmu:*
- 5: em estov:*
- 6: em estpl:*
- 7: em estro:*
- 8: em htc:*
- 9: gb est1:*
- 10: gb est2:*
- 11: gb htc:*
- 12: gb est3:*
- 13: gb est4:*
- 14: gb est5:*
- 15: em estfun:*
- 16: em estom:*
- 17: em gss hum:*
- 18: em gss inv:*
- 19: em gss pln:*
- 20: em gss vrt:*
- 21: em gss fun:*
- 22: em gss mam:*
- 23: em gss mus:*
- 24: em gss pro:*
- 25: em gss rod:*
- 26: em gss phg:*
- 27: em gss vrl:*
- 28: gb gssI:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3574.5	78.6	3163	11	AY109444 Zea mays
2	1518.5	33.4	2049	11	AY108805 Zea mays
3	1268.5	27.9	843	14	CB349719 CAB2SG000
4	1261.5	27.7	846	14	CB8944504 EST647296
5	1240.5	27.3	837	14	CB8944708 EST647270
6	1188.5	26.1	812	14	CB893989 EST646781
7	1188.5	26.1	818	10	BG648761 EST510380
8	1172	25.8	783	10	BE642455 Cri2.6.B1
9	1169.5	25.7	804	14	CB893763 EST646555
10	1152	25.3	740	13	BQ39239 WHE4163.H
11	1115	24.5	722	14	CB610803 ALBDO000
12	1115	24.5	750	28	BH603944 BOGTL30TF
13	1114.5	24.5	727	13	BQ993972 QGFSP11.Y
14	1111	24.4	736	10	BG648385 EST510004
15	1098.5	24.1	729	13	BQ862449 QGC21B24*
16	1098.5	24.1	747	10	BG647349
17	1097	24.1	783	14	CB893376 EST508968
18	1089	23.9	709	29	BZ963117 PUGFU64TD
19	1070	23.5	686	13	BQ869697 QGD6P04.Y
20	1054	23.2	753	29	BZ963112 PUGFU64TB
21	1044	22.9	863	28	BH465756 BOGFI04TF
22	1001.5	22.0	703	10	BF480849 L0-2608T3
23	991.5	21.8	818	28	BH250711 BOGAQ17TF
24	989	21.7	1100	29	CNS06844
25	983	21.6	1036	29	CNS06B91
26	969	21.3	900	28	BH163888 ENT5180TF
27	966.5	21.2	892	28	BH148894 ENTPE41TF
28	964	21.2	872	28	AZ671816 ENTXX12TR
29	957	21.0	895	28	AZ528202 ENTFC95TF
30	956	21.0	617	12	BJ291097 BJ291097
31	953	20.9	689	28	BH423861 BOGKS57TR
32	953	20.9	1798	11	AY103877 Zea mays
33	952.5	20.9	894	28	AZ679340 ENTU110TF
34	945	20.8	625	9	AW257917 867064B03
35	935	20.5	1015	28	AF106743 AF106743
36	933	20.5	903	28	BH132749 ENT0212TF
37	930	20.4	863	28	AZ679319 ENTJ55TR
38	928	20.4	935	14	CB894408 EST647200
39	926.5	20.4	813	28	AZ528443 ENT0M71TR
40	926	20.4	598	14	CB894119 EST646911
41	923.5	20.3	890	28	AZ548311 ENTGD36TR
42	923	20.3	621	13	BH892034 P058E06.P
43	918.5	20.2	916	28	BH164525 ENT2D3TR
44	917	20.2	692	10	AW930379 EST340932
45	915.5	20.1	947	28	BH130876 ENTPEB12TF

ALIGNMENTS

RESULT	1	3163 bp	mRNA	linear	HTC	17-OCT-2002
LOCUS	AY109444	Zea mays	CU1518.1	mRNA sequence.		
DEFINITION	Zea mays					
ACCESSION	AY109444					
VERSION	AY109444.1	GI:21213165				
KEYWORDS	HTC.					
SOURCE	Zea mays					
ORGANISM	Zea mays					
REFERENCE	1	(bases 1 to 3163)				

AUTHORS

Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whittitt, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M., and Tingey, S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 3163)
Coe, E.H.

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA

If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSU, maizep.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.

FEATURES

source

Location/Qualifiers

1..3163
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:630374"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

626 a 846 c 942 g 490 t 259 others

SEQUENCE

ALIGNMENT

SCORES

ad. No.:

Length:

Matches:

Conservative:

Mismatch:

Indels:

Gaps:

1 MetAspProGluysPheThrHisLysThrAsnGluThrIleAlaThrAlaHisGluLeu 20

154 ATCAATCCGACAACTTACCACAGACGAGCGGCGATCGTNNNNCGACGAGATT 213

21 AlaValAsnAlaGlyHisAlaGlnPheThrProLeuHisLeuAlaGlyAlaLeuSer 40

214 GCGGTGGAGCGCGCCACGCGGAGTCCAGCGGCTGACCTGGCGCGAGTGTGGCTGG 273

41 AspProThrGlyIlePheProGlnAlaIleSerSerAlaGlyGlyGluAsnAlaAlaGln 60

274 GACAAGGGCGGATCTCGGCGAGGCGCATCNNNNNNNNNNNNNNNNNNNNNNNNN 333

61 SerAla-----GluArgValIleAsnGlnAlaLeuLysLysLeuProSerGlnSerPro 78

334 NNNNNNNCGTTCGAGCGCGTGTGAACAACTCGCTCAAGAACCTCGCGTCCGCGCCG 393

79 ProProAspAspIleProAlaSerSerSerLeuLysValIleArgArgAlaGlnAla 98

394 CCGCGGATCTCGCGCGGCTCCAGCGGCTGATCAAGGTTCATCCGCGGCGCGAGTCC 453

99 AlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuIleMetGlyLeuLeu 118

454 GCGGAGAAAGCGGGGACTCGACCTCGCGCTCGACCGAGTCTGCTCGCGCTGCTC 513

119 GluAspSerGlnIleArgAspLeuLeuAsnGluValGlyValAlaThrAlaArgValLys 138

514 GAGGACTCGAGATCTCCGACTGCTCAAGGNNNNNNNNNNNNNNNNNNNNNNNNNN 573

139 SerGluValGluLysLeuArgGlyLysGluGlyLysValGluSerAlaSerGlyAsp 158

574 NNNNGCTTGGAGNN 633
159 ThrAsnPheGlnAlaLeuLysThrTyrGlyArgAspLeuValGluGlnAlaGlyLysLeu 178
634 ACCAACTTCCAGCGCTCAAGACATACCGCGGACCTCGTCGAGCAGCGCGGAGCTG 693
179 AspProValIleGlyArgAspGluLeuArgValValArgValValArgValValArg 198
694 GACCCCGCTCATCGCGCGGACGAGGATCGCGCGCTCGTGGCATCTCTCGCGCGC 753
199 ThrLysAsnAspProValLeuIleGlyGluProGlyValGlyLysThrAlaValGlu 218
754 ACCAAGAACACCCCGCTCTCATCGCGAGCCCGCTTGGCAAGACGCGCTCGTGAG 813
219 GlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspValArg 238
814 GGCCTCGCGAGCGCATCTTTCGCGGAGCTGCCAGTAACCTCTCGAGCTCGGCTC 873
239 IleSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGlu 258
874 ATCGCGCTCGACANNN 933
259 ArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeuPhe 278
934 CGGCTCAAGCGCTGCTCAAGAGGTGAGAGGCGCGGAGGAGGATCTTCTTCTCATC 993
279 AspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAlaAsn 298
994 GACGAGATACACCTCGTCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1053
299 LeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrLeuGlu 318
1054 CTGTTCAAGCCATGCTGGCGAGGAGCAGCTCAGGTGCTATCGCGCCACCTCGAG 1113
319 GluTyrArgLysTyrValGluLysAspAlaAlaPheGluArgArgPheGlnGlnVal 338
1114 GAGTACCGCAAGTACGTGGAGAGGACGACGCTTCAGCGCGGCTTCAGCAGGTGTT 1173
339 ValAlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGlyTyr 358
1174 GTCGCGAGCGCGAGCGTGCACGCTCAGCATCTGAGGGGACTCAAGGAGAGTATC 1233
359 GluGlyHisGlyValArgIleGlnAspArgAlaLeuLeuAlaGlnLeuSer 378
1234 GAGGCGCACCATGCTGGCGAGGATCCAGACCGCGCTCTGTTGTCGCGGACAGCTCT 1293
379 AlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGlu 398
1294 GCGAGGTACATCATGCTCGCGACCTGCTGACAAAGCCATACCTGTTGAGCAGGCG 1353
399 CysAlaAsnValArgValGlnLeuAspSerGlnProGluIleAspAsnLeuGluArg 418
1354 TCGGCCAATGTAGGGTGCAGCTCGACAGCAGCGAGGAGATTGATTAACCTGGAGAG 1413
419 LysArgMetGlnLeuGluLeuHisAlaLeuGluArgGlyLysAspLysAlaSer 438
1414 AAGAGATCCAGTTGAGTTGAGTTCACCGCTCGAAGGAGGAGGAGGAGGAGGAG 1473
439 LysAlaArgLeuIleGluValArgLysGluLysGluLysAspLeuArgAspLysLeu 458
1474 AAAGCCCGCTGATGAGTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1533
459 LeuThrMetLysTyrArgLysGluLysGluArgIleAspGluIleArgGluLeuGln 478
1534 CTGACCATGAAGTANN 1593
479 LysArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAlaArg 498
1594 CGCGCGAGGAGTTCAGTTCCACCTGCGAGGAGGCGCGCGGAGGAGGAGGAGGAG 1653
499 AlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 518

1654 GTGGCGGACCTCAAGTACGGCCCTCCAGGAATCGAGCGCGCGATCTCAAGCTN--- 1710
 519 GlyThrSerSerGluuAenValMetLeuThrGluuAenValGlyProGluHisIleAla 538
 1711 NNNNNNNNNNNNNNNNNNNNTGATGCTCACCGAAACCGTCGGCCCTGAACAATTGCA 1770
 539 GluValValSerArgTyrThrChlyIleProValThrArgLeuGlyGlnAenGluysGlu 558
 1771 GAGGTGGTGAGCGGTGGACGGGTATTCCAGTGACCGCGCTTGGCGCAAGACCAAGGAG 1830
 559 ArgLeuIleGlyLeuAlaAspArgLeuHisArgValValGlyGlnAenGluAlaVal 578
 1831 AGGCTGGTGGCTGGCTGGCTGACAGGCTTACAGAGGGTGGTCGGCCACAGAGGCTGTG 1890
 579 AenAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnPro 598
 1891 AGCGCGCTCGAGAGCGGCTGAGGTGAGGGCGGTCTTGGCAGGCCACACAGGCC 1950
 599 ThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaIleAla 618
 1951 ACTGCTCGTTCCTCTCTGGTCCGACTGGCGTGGGGAACCTGAGCTGGCCAGGCC 2010
 619 LeuAlaGluGlnLeuPheAspAspGluAenLeuValArgIleAspMetSerGluTyr 638
 2011 CTAGCCGAACAGCTGTTCGACGACGAGAACCTTCTTCCGATCGACATGTCGAGTAC 2070
 639 MetGluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHisGlu 658
 2071 ATGAGCAGCAGCTCGGTTCCTCGCTCATCGAGCACCCACCTGGCTACGTCGGCATGAA 2130
 659 GluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAsp 678
 2131 GAGGTGGCGAGCTGACAGCAAGTAGGAGGAGGCGGTACACGCGTATCTGTTCCGAC 2190
 679 GluValGluLysAlaHisValAlaValPheAenThrLeuLeuGlnValLeuAspAspGly 698
 2191 GAGTTCGAGAAGGCCATGTCGCGGTGTTCAACACCTGCTCCAGGCTTCGACGAGCGC 2250
 699 ArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAenSerValIleMetThr 718
 2251 AGGTGACGATGGCGAAGCGGAGCGGAGCTTCAGGAACACCGGTATCATGATGACA 2310
 719 SerAenLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluVal 738
 2311 TCGAACTCGCGCGCGAGCACCTCTCGTGAGATGGTGGCGAGAACTCCATGAAGGTC 2370
 739 AlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAenArg 758
 2371 GCTCGCATCTGGTTCATGCGAGAGGTGAGGAGGCACATTCGCGCTGAGCTGCTGAACGT 2430
 759 LeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAlaArg 778
 2431 CTGACAGATCGTGATCTTCGATCTCTGCTCCACGAGCAGCTGAGGAAGGTGCTCGC 2490
 779 LeuGlnMetLysAspValAlaValArgLeuAlaGluArgGlyValAlaLeuAlaValThr 798
 2491 CTTGATGAGAGATGGCGGTCGCTTCGCGAAGGGGATCGCTCTGGCTGTGACC 2550
 799 AspAlaAlaLeuAspTyrIleLeuAlaGluSerTyrAspProValTyrGlyAlaArgPro 818
 2551 GAGCGCATTTGACATCATCTGCTCTCTTACGATCGGTGATCGCGCGCGGCCA 2610
 819 IleArgArgTyrMetGluLysLysValValThrGluLeuSerLysMetValValArgGlu 838
 2611 ATCAGGAGGTGGATCGAGAAGAGGGTGGTGCAGCGCTGCGAAGATGCTGATCCAGGAG 2670
 839 GluIleAspGluAenSerThrValTyrIleAspAlaGlyAlaGly-----AspLeuVal 856
 2671 GAGTTCAGCAGAACTCCAGGCTTACATCGACCGCGCGCGCGAAGCAGCGGTGTC 2730
 857 TyrArgValGlu---SerGlyGlyLeuValAspAlaSerThrGlyLysLysSerAspVal 875
 2731 TACAGGTGACCGGCGCGGTCTGTTGTAACGCTCAGACGCGGGATGAGTCTGGACATC 2790

876 LeuIleHisIleAlaAenGlyProLysArgSerAspAlaAlaGlnAlaValLysMet 895
 2791 CTGATCAGGTCCCAACAGCTCCACCGAGGAGCGCTGCGCAGCGCGTCAAGAAGATG 2850
 896 ArgIle---GluGluIleGluAspAspAspAsnGluGlu 907
 2851 AGGATCATGAGGAGGAGGAGGAGCGGATGACGAGGAG 2889
 RESULT 2
 AY108805
 LOCUS Zea mays PC0077883 mRNA sequence.
 DEFINITION Zea mays PC0077883 mRNA sequence.
 ACCESSION AY108805
 VERSION AY108805.1 GI:21212055
 KEYWORDS HTC.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 2049)
 AUTHORS Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
 Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
 TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
 Overgo Probes
 JOURNAL Unpublished (2002)
 REFERENCE 2 (bases 1 to 2049)
 AUTHORS Coe, E.H.
 TITLE Direct Submission
 JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA
 COMMENT If you are interested in getting corresponding physical clones,
 these are publicly available from ZmDB and may be found by BLAST
 searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
 www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
 maize cDNA sequences is either Virginia Walbot, Stanford or Pat
 Schnable, Iowa State, then clones may be requested from ZmDB:
 www.zmdb.iastate.edu.
 FEATURES
 Location/Qualifiers
 1..2049
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="MaizeDB:634307"
 /db_xref="taxon:4577"
 /clone_lib="Maize Mapping Project/DuPont Consensus
 Library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed DuPont contigs; this resource was
 assembled by DuPont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"
 BASE COUNT 615 a 409 c 565 g 460 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,928-134 Length: 2049
 Score: 1518.50 Matches: 318
 Percent Similarity: 65.99% Conservative: 138
 Best Local Similarity: 46.02% Mismatches: 166
 Query Match: 33.37% Indels: 70
 DB: 11 Gaps: 8
 US-09-812-350-17 (1-911) x AY108805 (1-2049)
 200 LysAenAenProValLeuIleGlyGluProGlyValGlyLysThrAlaValValGluGly 219
 14 AAGAACAATCCCTGTTGATTGGAGAGCGCTGCTGGAAAAAAGCTGCAATTGCTGAAGGA 73
 220 LeuAlaGlnArgIleValLysGlyAspValProAenSerLeuThrAspValArgLeuIle 239
 74 CTTGCTCAGCGCATTTCTTACAGGCGATGTACCTGAAACAATAGAGGAAAAAGGTGTCATA 133

240 SerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyrArgGlyGluPheGluGluArg 259
 134 ACCCTTGACATGGGACTCTCTGTTGCTGGCACAAGTACCGTGAGAAATTCGAAGAAGA 193
 260 LeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValLeuPheLeuAsp 279
 194 TTAAGAAGCTGTGAGGAGAAATAAAGCAAGTAT---GAGATAATCTCTTTATTGAT 250
 280 GluLeuHisLeuValLeuGlyAlaGlyLysThrCluGlySerMetAspAlaLeuLeu 299
 251 GAAGTTCACACTCTGATAGGAGCAGGAGC-AGCGAGGTGCTATAGATGCTGTAATATC 309
 300 PheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThrLeuGluGlu 319
 310 TTGAAGCCTGCTGGCAGAGGTGAATACAGTGCATTCGAGCCACTACACTAGATGA 369
 320 TyrArgLysThrValGluLysAspAlaPheGluArgPheGlnGlnValTyrVal 339
 370 TATAGGAAGCACATTTGAGAAGACCCAGCAGCTTGAACGAGGTTTCAACCTGTGAAGTG 429
 340 AlaGluProSerValProAspThrIleSerIleLeuArgGlyLeuLysGluLysTyrGlu 359
 430 CCAGAACCAACAGTAGATGAACCATTTGAATCCTCAGAGACTGAGGGAACCATATGAG 489
 360 GlyHisGlyValArgIleGlnAspArgAlaLeuIleAsnAlaAlaGlnLeuSerAla 379
 490 ATCCACCATAACTCTGTTACACTGATGAAGCTCTGATTCAGCTCAAAAGCTGTCTAT 549
 380 ArgTyrIleThrGlyArgHisLeuProAspLysAlaIleAspLeuValAspGluAlaCys 399
 550 CAATATATCATGATCGGTTCTCCAGATTAAGCAATTTGACTTGAATGATGAAGCAGGT 609
 400 AlaAsnValArgValGlnLeuAspSerGlnProGluGluIleAspAsnLeuGluArgLys 419
 610 TCCCGTGTAGGTACAGCATGCACAGTCCCGAGAGCAAGAGAGAGAGCTTGCAGAG--- 666
 420 ArgMetGlnLeuGluIleGluLeuHisAlaLeuGluArgGlyLysAspLysAlaSerLys 439
 667 -----GAGCTCAAAAGTCAACGAAACAGAGAAGTCAAGCTGTGCA 708
 440 AlaArgLeuIleGluValArgLysGluLeuAspAspLeuArgAspLysLeuGlnProLeu 459
 709 AGCCAGGATTTGAGAAGGCTGGGGAATTTGAGAGC----- 744
 460 ThrMetLysThrArgLysGluLysGluArgIleAspGluIleArgArgLeuLysGluLys 479
 745 -----CGTGAATGGAAATTGAAGGCCAGATTAACAGCCCTCATTTGACAG 789
 480 ArgGluGluLeuMetPheSerLeuGlnGluAlaGluArgArgTyrAspLeuAlaArgAla 499
 790 AGCAAGGAATTG----- 801
 500 AlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGly 519
 802 -----AGCAAGCAGAGAA 816
 520 ThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGlu 539
 817 GAGTCTGGAGAGACAGGACTATGGTCAATGAA-----GAGATATCCAGCAC 864
 540 ValValSerArgTyrThrGlyLeuProValThrArgLeuGluGlnAsnGluLysGluArg 559
 865 ATAGTATCTTCATGGACTGGCATCCCTGTGGAGAGGTTTCCAGGATGAATCTGATAAG 924
 560 LeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGluAsnGlnAlaLeuAsn 579
 925 CTCTTAAGATGGAAGAGACTTTGCAACGCGTGTCAATGGCCCAAGATGAGGCTGTGTA 984
 580 AlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThr 599
 985 GCAATTAGTCGTCCTCCGCGTCTGCTGGTGGGTCTCAAGAACCCCAAGGCAATTT 1044

QY 600 GlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeu 619
 Db 1045 GCAAGCTTTATTTTGTGGTCCACCGCGCTGGGAAGTCTGAGCTTGCAGAGGCTCTT 1104
 QY 620 AlaGluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluTyrMet 639
 Db 1105 GCAGCCTATTACTTTGGCTCTGAGGAGGCTATGATCCGCTTGATATGATGATTCATG 1164
 QY 640 GluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHisGluGlu 659
 Db 1165 GAGAGACACACGATTCACAGCTATTCGTTTCCCTCCAGGATATGATGATGATGATGATG 1224
 QY 660 GlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPheAspGlu 679
 Db 1225 GGTGGCCAGCTGACACAGGAGGCTTCGAGCGGCGCATACACAGTTGTGCTCTTTGATGAG 1284
 QY 680 ValGluLysAlaHisValAlaValPheAsnThrLeuGlnValLeuAspAspGlyArg 699
 Db 1285 ATTGAAGAGGACACACCTCTGATGCTCTTCAACATGATGCTTCAAGATGGGAGA 1344
 QY 700 LeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleLeuMetThrSer 719
 Db 1345 TTGACTGACACAAAGGAGGAGCGTGGACTTCAAGAACACACTCTCTGATCATGACCTCA 1404
 QY 720 AsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAla 739
 Db 1405 AATGTAGGAGCAGCTGCTCATCGAAGGCT---GGAAGGAGATTCGATTTGACCTTGAC 1461
 QY 740 ArgAsp-----CysValMetArgGluValArg 748
 Db 1462 TCTGATGAAGAGGACAGTAGCTACAGCAGGATCAAGAGCTGGTTCATCGAGGAGATGAAG 1521
 QY 749 LysHisPheArgProGluLeuLeuAsnArgLeuAspGluIleValValPheAspProLeu 768
 Db 1522 CAGTATTTCCGACCTGAGTTCTCTCAACCGTCTCGATGAGATGATCGTGTTCAGCAGCTT 1581
 QY 769 SerHisAspGlnLeuArgLysValAlaArgLeuGlnMetLysAspValAlaValArgLeu 788
 Db 1592 ACCAGCTCGAGGTCAGGAGATAGCGGACATCATGCTCCAGGAGGCTCTTGACAGGCTG 1641
 QY 789 AlaGluArgGlyValAlaLeuAlaValThrAspAlaAlaLeuAspTyrIleLeuAlaGlu 808
 Db 1642 AAGGCCAAGGACATCAATCTTCAAGTACCCGAGAGTTCAAGGAGCGGCTGGTGGAGAA 1701
 QY 809 SerTyrAspProValTyrGlyAlaArgProIleArgArgTrpMetGluLysLysValVal 828
 Db 1702 GGCTACACACCTAGCTATGTTGTCACGCCCTGAGGCGGACCATCATGAGGCTGCTGGAG 1761
 QY 829 ThrGluLeuSerLysMetValValArgGluGluIleAspGluAsnSerThrValTyrIle 848
 Db 1762 GACAGCCTTCTGAGAGATGCTCGCAGGAGGAGTAAGAGGCGGACCTTGCCATAGTA 1821
 QY 849 AspAlaGlyAla---GlyAspLeuValTyrArgValGluSerGlyGlyLeuValAspAla 867
 Db 1822 GATGTGACCTCGAGGAGGAGGTTGTTGCTCAATGCTCAGGCGGCGCATACCGGAGCTC 1881
 QY 868 SerThrGlyLysLysSerAspValLeuIleHis 878
 Db 1882 TCAACTCCGCGGATCCACCTTTAGCTCTGTACAT 1914

RESULT 3
 CB349719 843 bp mRNA linear EST 10-APR-2003
 LOCUS CAB2SG0007 IIIaf_A11 Cabernet Sauvignon Berry - CAB2SG Vitis
 DEFINITION Vitis vinifera cDNA clone CAB2SG0007_IIIaf_A11 5', mRNA sequence.
 ACCESSION CB349719
 VERSION CB349719.1 GI:28970686
 KEYWORDS EST.
 SOURCE Vitis vinifera
 ORGANISM Vitis vinifera
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; Vitaceae; Vitis.

1 (bases 1 to 843)
 Goes da Silva, F., Iandolo, A., Lim, H., Baek, J., Jones, K. and Cook
 D.
 Expressed sequence tags from cabernet sauvignon berries at various
 developmental stages
 Unpublished
 Contact: Douglas Cook, PhD
 CAES Genome Facility
 UC Davis, Plant Pathology
 One Shields Ave, Davis, CA 95616, USA
 Tel: 530 754 6561
 Fax: 530 754 6617
 Email: drcook@ucdavis.edu
 Seq primer: ACGGTACCGACATATGCC.
 Location/Qualifiers
 1..843
 /organism="Vitis vinifera"
 /mol_type="mRNA"
 /cultivar="Cabernet Sauvignon"
 /db_xref="taxon:29760"
 /clone="CAB2SG0007.IIIaF_A11"
 /sex="Hermaphrodite"
 /dev_stage="Veraison"
 /lab_host="DHSalpa"
 /clone_lib="Cabernet Sauvignon Berry - CAB2SG"
 /note="Organ: Berry; Vector: pDNR; Site 1: Sfil; Site 2:
 Sfil; CAB2SG is a cDNA library of Cabernet Sauvignon Clone
 8 berries. Samples were collected at veraison (transition
 between stage II and stage III of berry growth) from
 field-grown vines 60 days after full bloom. Berries were
 still green but soft. Sampled vines were located at the
 University of California, Davis, experimental vineyard.
 cDNAs were made by oligo-dT priming and directionally
 cloned. 5' and 3' adaptors were used in cloning as follows:
 5'-AAGCAGTGTATCAAGCAGAGTGCCTACGCGGG-3' and
 5'-ATTGAGGCGGAGCGGCGGACATG-dT(30)NN-3'. Library was
 constructed using the Clontech Creator SMART kit and
 size-selected to contain the 0.5-3 kb size fraction."

FEATURES
 source

E COUNT 218 a 166 c 251 g 207 t
 GIN

gment Scores:

d. NO.: 5,238-111 Length: 843
 re: 1268.50 Matches: 250
 cent Similarity: 95.73% Conservative: 19
 t Local Similarity: 88.97% Mismatches: 11
 ry Match: 27.88% Indels: 1
 14 Gaps: 1

09-812-350-17 (1-911) x CB349719 (1-843)

498 ArgAlaAlaAspLeuArgTyrGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeu 517
 2 AGAGTGTGCTGATTGGCATATGGGCAATTCAGAGAGTGAGGCTGTACTAGCAACCTT 61
 518 GluGlyThrSerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIle 537
 62 GAAGGAACCC---ACTGATGAGAACAATGATGTTAAACGGAGACTGTTGGGCTGAGCAGATT 118
 538 AlaGluValValSerArgTyrThrGlyIleProValThrArgLeuGlyGlnAsnGluIys 557
 119 GCTGAGGTGGTGAGCGCTGCTGGCATCCCTGTCTAGCTTGGACAAATGATATAA 178
 558 GluArgLeuIleGlyLeuAlaAspArgLeuHisIysArgValValGlyGlnAsnGlnAla 577
 179 GAAAGTTTAAATGGGCTTGGGAGAGGCTGCATCAGAGAGTGGTGGACAGGATCAGGCA 238
 578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGln 597
 239 GTGAGTGTGTGGCAGAGGCTGCTTGTAGATCGAGGGCTGGGTTAGGAGGCCCAACAA 298
 598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyValThrGluLeuAlaIys 617

Db 299 CCCACTGGTTCGTTCTCTCTTGGTGGTCCAACTGGTGTGTGGTAAGACTGAGCTTGCCAG 358
 Qy 618 AlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValArgIleAspMetSerGlu 637
 Db 359 GCCCTGGCTCAAGACGCTCTTTGATGATGAAAAATCTTTGGTGGAATTTGACATGTCCGAG 418
 Qy 638 TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProGlyTyrValGlyHis 657
 Db 419 TATATGAGCAGCAGCATTCAGTTTCAAGATTGATTTGGGCTCTCTCTGGATATGTTGACAT 478
 Qy 658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe 677
 Db 479 GATGAGGTGGGCAACTTCACAGAGGAGTAAGAGGAGGCTTATAGTGTGTACTGTTTC 538
 Qy 678 AspGluValGluIysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspAsp 697
 Db 539 GATGAGGTAGAAAGACATATCGCGGTATTCATACCTCTCTCAGGTTTATAGACGAT 598
 Qy 698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMet 717
 Db 599 GGAAGGTAACTGATGGCCAGGCGGCACCTGTAGACTTCACAAACACACATGATCATATG 658
 Qy 718 ThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGlu 737
 Db 659 ACTTCAACCTTGGAGCTGAGCATCTCTCTCGGGTTTAGTAGGCAAGTGTACATGCAA 718
 Qy 738 ValAlaArgAspCysValMetArgGluValArgLysHisPheArgProGluLeuLeuAsn 757
 Db 719 GATGCTGTACCGGTGTATGTCAGGAGGTGAGNAGGCACCTTCAGGCCAGAGTTGCTTAAC 778
 Qy 758 ArgLeuAspGluIleValValPheAspProLeuSerHisAspGlnLeuArgLysValAla 777
 Db 779 CGGTGATGAATTTGGTATTGATCCCTCTCATCATGATCACTGAGAAAGGTTGCC 838
 Qy 778 Arg 778
 Db 839 AGG 841

RESULT 4
 LOCUS
 DEFINITION

CB894504 846 bp mRNA linear EST 24-APR-2003
 EST647296 HOGA Medicago truncatula cDNA clone HOGA-31M8, mRNA
 sequence.

ACCESSION
 VERSION

CB894504
 CB894504.1 GI:30101673

KEYWORDS

EST.
 Medicago truncatula (barrel medic)

SOURCE

ORGANISM

Medicago
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.

REFERENCE

1 (bases 1 to 846)
 Hahn, M.G., Ojanen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S.,

Utterback, T., Cho, J. and Fraser, C.M.

ESTs from roots of Medicago truncatula treated with

oligolacturonides of DP 6-20

Unpublished

JOURNAL

COMMENT

Contact: Michael G. Hahn

Complex Carbohydrate Research Center

University of Georgia

220 Riverbend Road, Athens, GA 30602-4712, USA

Tel: 706-542-4457

Fax: 706-542-4412

Email: hahn@ccrc.uga.edu

TIGR sequence name: MTMDT76TK

More information is available at: www.medicago.org

Seq primer: SKmod (CTA GAA CTA gtg gat CC).

FEATURES

Location/Qualifiers

1..846

/organism="Medicago truncatula"

/mol_type="mRNA"

/cultivar="Al7"


```

/db_xref="taxon:3880"
/clone="HOGA-31M8"
/tissue_type="3 day old seedling roots"
/dev_stage="24 hours after treatment in the dark at 26 C
with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
presence of 100 ug/ml Gentamicin"
/lab_host="XJOLR"
/clone_lib="HOGA"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in SOLR cells."
BASE COUNT      270 a 139 c 243 g 194 t
UGIN

Alignment Scores:
ed. No.:      2,45e-110      Length:      846
Matches:      1261.50
Conservative:  95.04%
Mismatch:      88.30%
Indels:        27.73%
Gaps:          14

389 AspLysAlaIleAspLeuValAspGluAlaCysAlaAsnValArgValGlnLeuAspSer 408
1 GACAGGCCATTGATTAGTTGATGGCATGTCTAATCTGAGAGTTCACTTGATAGT 60
409 GlnProGluLuleAspAsnLeuGluArgLysArgMetGlnLeuGluLeuHis 428
61 CAACCTGAGCAATTCACATCTTGAAGAGAGAGAAATGCAATTAGAGTTGAATGTCAT 120
429 AlaLeuGluArgGluLysAspLysAlaSerLysAlaArgLeuLeuGluValArgLysGlu 448
121 GCTCTAGAGAAAGAGAGAGAGCAAGCTAGCAAGCCCGCTCTGTAGATGTACGAGAGAA 180
449 LeuAspAspLeuArgAspLysLeuGlnProLeuThrMetLysTyrArgLysGluLysGlu 468
181 CTTGATGACCTTGAGAGACAGCTTCAACCTCTGAGATGAAGTATAGTAAGAGAGAGAG 240
469 ArgIleAspGluLuleArgLysGlnLysGlnLysArgGluLuleuMetPheSerLeuGln 488
241 AGGATTGATGAGATTCGAGGCTGAAGCAGCAACACGTGAGAGAGCTCTCTTTGCACTACAG 300
489 GluAlaGluArgTyrArgLeuAlaAlaGlnLeuGluGlyThr-----SerSerGluGluAsn 508
301 GAGCAGAGAGCGGTATGATCTAGCAGAGCTGCGGACCTGCGATGATGTCGATGAA 360
509 GluValGluSerAlaIleAlaGlnLeuGluGlyThr-----SerSerGluGluAsn 525
361 GAGTGGAACCTGCAATTAATAATCTTGAAGGTAGCACTGATGGGAACACGGATGAGAAC 420
526 ValMetLeuThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTyrThr 545
421 TTAATGTGTCGGAACACAGTGTGACCGGACCAATAGCCGAGGTGTGTGCCGATGGACT 480
546 GlyIleProValThrArgLeuGluGlnAsnGluLysGluArgLeuLeuGlyLeuAlaAsp 565
481 GGTATACAGTACTAGCTCGGCCCAAAATGAGAAAGCAAGGTGTGTGACCTGGTGAC 540
566 ArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIle 585
541 AGACTGCACACTAGAGTCGTGGGACAGACCAAGCGGTAAATGCTGTGCTGAGGCTGTA 600
586 LeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeu 605
601 TTGAGATCAAGAGCTGTTTAGGAGAGACCCCAACCAACTGCTCTTCTTATTCCTT 660
606 GlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluLynLeuPheAsp 625

```

```

661 GGTCCGACTGCTGTTGGTAAACCTGAGCTTGCAAGGCTCTTGCAGAGCAGCTATTGAT 720
626 AspGluAsnLeuLeuValArgLysMetSerGluTyrMetGluGlnHisSerValSer 645
721 GATGAAATCATCTGTTGGTGAATTCATGTCGGAATACATGGAACACACTCTGTATCA 780
646 ArgLeuIleGlyAlaProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGlu 665
781 AGATTGATTGTCACACACAGGCTATGTCGATCAGGAGGAGGGTCAATTAAACAGAA 840
666 AlaVal 667
841 GCTGTA 846

RESULT 5
CB894478
LOCUS      EST647270 HOGA Medicago truncatula cDNA clone HOGA-31G20, mRNA
DEFINITION
ACCESSION  CB894478
VERSION    CB894478.1
KEYWORDS   EST.
SOURCE     Medicago truncatula (barrel medic)
ORGANISM   Medicago truncatula
REFERENCE  1 (bases 1 to 837)
AUTHORS    Hahn, M.G., Ojansen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S.,
            Utterback, T., Cho, J., and Fraser, C.M.
TITLE      ESTs from roots of Medicago truncatula treated with
            oligogalacturonides of DP 6-20
JOURNAL    Unpublished
COMMENT    Contact: Michael G. Hahn
            Complex Carbohydrate Research Center
            University of Georgia
            220 Riverbend Road, Athens, GA 30602-4712, USA
            Tel: 706-542-4457
            Fax: 706-542-4412
            Email: hahn@ccrc.uga.edu
            TIGR sequence name: MWMDT46TK
            More information is available at: www.medicago.org
            Seq primer: SKmod (CTA GAA CTA gTg gAT CC).
            Location/Qualifiers
                1..837
                    /organism="Medicago truncatula"
                    /mol_type="mRNA"
                    /cultivar="Al7"
                    /db_xref="taxon:3880"
                    /clone="HOGA-31G20"
                    /tissue_type="3 day old seedling roots"
                    /dev_stage="24 hours after treatment in the dark at 26 C
                    with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
                    presence of 100 ug/ml Gentamicin"
                    /lab_host="XJOLR"
                    /clone_lib="HOGA"
                    /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
                    XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                    was directionally ligated into the Unizap XR vector from
                    Stratagene and packaged using Gigapack III Gold packaging
                    extracts. Plasmids containing cDNA inserts were excised
                    from the recombinant lambda-Zap phage using Ex-assist
                    helper phage and propagated in SOLR cells."
BASE COUNT      258 a      147 c      236 g      196 t
ORIGIN

Alignment Scores:
Pred. No.:      2,47e-108      Length:      837
Score:          1240.50
Percent Similarity: 94.24%
Best Local Similarity: 87.77%
Matches:        1244
Conservative:    244
Mismatch:        18

```

٥١

•

40

५।

15

10.

35

AA

191

—

1

7.7

27

ICT

ily

二七

•

301 AGCACTGATGGGACACCGATGAGCACTTATGTTGACGGAAACAGTTGGACCGACCAA 360
 537 IleAlaGluValSerArgTrpThrGlyIleProValThrArgLeuGlyGlnAsnGlu 556
 361 ATAGCCGAGGTGTTAGCCGATGAGCTGGTATACAGTACTAGCTCGGCCAAATGAG 420
 557 LysGluArgLeuIleGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGln 576
 421 AAAGCAAGGTGGTGGGACTTGGTGACAGACTGCACACTAGAGTCGTGGGACAGACCAA 480
 577 AlaValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyValGln 596
 481 GCGGTATATGCTGTGCTGAGCTGATTTGATGATGAAATCAGTTGGTGAGAAATGACATGCG 540
 597 GlnProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAla 616
 541 CAACCACTGCTGCTCTTATTCCTTGGTCCGACTGGTGTGGTAAACCTGAGCTTGCA 600
 617 LysAlaLeuAlaGluGlnLeuPheAspAspGluAsnLeuLeuValAlaGlyIleAspMetSer 636
 601 AAGGCTCTTGACAGCGAGCTATTTGATGATGAAATCAGTTGGTGAGAAATGACATGCG 660
 637 GluTyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGly 656
 661 GAATACATGGAACACACTCTGTATCAGATTGATTGGTGACACACAGGGTATTTGGA 720
 657 HisGluGlyGlyGlnLeuThrGluAlaValArgArgProGlyCysValIleLeu 676
 721 CATGAGGAAGGGGCTCAATTAACAGAGCTGTAGGAGAGAGGCCATACAGTGTGTACTC 780
 677 PheAspGluValGluLysAlaHisValAla 686
 781 TTGATGAAGTTGAGAAAGCACACATCT 810

SULT 7
 648761
 CUS
 EST510380 HOGA Medicago truncatula cDNA clone phOGA-23B24 5' end,
 mRNA sequence.
 BG648761
 BG648761.1 GI:13783873
 EST.
 Medicago truncatula (barrel medic)
 Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 1 (bases 1 to 818)
 Hahn, M.G., Ojanen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S.,
 Utterback, T., Cho, J. and Fraser, C.M.
 ESTs from roots of Medicago truncatula treated with
 oligogalacturonides of DP 6-20
 Unpublished
 Contact: Michael G. Hahn
 Complex Carbohydrate Research Center
 University of Georgia
 220 Riverbend Road, Athens, GA 30602-4712, USA
 Tel: 706-542-4457
 Fax: 706-542-4412
 Email: hahn@ccrc.uga.edu
 G392367e TIGR sequence name: MTWCP12TK More information is
 available at: www.medicago.org
 Seq primer: SKmod (CTA GAA CTA gtg GAT CC).
 Location/Qualifiers
 1. .818
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="AL7"
 /db_xref="taxon:3880"
 /clone="phOGA-23B24"
 /tissue type="3 day old seedling roots"
 /dev_stage="24 hours after treatment in the dark at 26 C

with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
 presence of 100 ug/ml Gentamicin"
 /lab_host="XLOLR"
 /clone_lib="HOGA"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
 was directionally ligated into the Unizap XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-zap phage using Ex-assist
 helper phage and propagated in SOLR cells."

BASE COUNT 239 a 151 c 230 g 198 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2,27e-103 Length: 818
 Score: 1188.50 Matches: 232
 Percent Similarity: 93.75% Conservative: 23
 Best Local Similarity: 85.29% Mismatches: 14
 Query Match: 26.12% Indels: 3
 DB: 10 Gaps: 1

US-09-812-350-17 (1-911) x BG648761 (1-818)

QY 484 MetPheSerLeuGlnGluAlaGluArgTrpArgLeuAlaArgAlaAspLeuArg 503
 Db 2 CTCCTTGCACCTACAGAGGCGAGAGGCGGTATGATCTAGCAGAGCTGCGGACTGCGA 61
 QY 504 TGTGlyAlaIleGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThr----- 520
 Db 62 TATGTCGAATTTGAAGAGGTGGAACCTCAATTTAAATCTTTGAAGGTAGCCTCATGGG 121
 QY 521 SerSerGluGluAsnValMetLeuThrGluAsnValGlyProGluHisIleAlaGluVal 540
 Db 122 AACCGGATGAGACTTAATGTTGAGGAAACAGTTGGACCGCAACCAATAGCGAGGTT 181
 QY 541 ValSerArgTrpThrGlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeu 560
 Db 182 GTTACCCGATGAGCTGTGTATACAGTGTAGCTCGGCCCAAAATGAGAAAGCAAGGTTG 241
 QY 561 IleGlyLeuAlaAspArgLeuHisArgValValGlyGlnAsnGlnAlaValAsnAla 580
 Db 242 GTTGGACTGTGTGTCAGACTGCACACTAGATGTGTGGACAGCAAGCGGTTAATGCT 301
 QY 581 ValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGly 600
 Db 302 GTTGTCTGAGGCTGTATTGAGATCAAGAGCTGGTTAGGAGACCCCAACCACTAGTGT 361
 QY 601 SerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAla 620
 Db 362 TCCTTCTTATTCCTTGGTCCGACTGTGTGGTAAACTGAGCTTGCAGAGGCTCTTGCA 421
 QY 621 GluGlnLeuPheAspAspGluAsnLeuValArgIleAspMetSerGluTyrMetGlu 640
 Db 422 GAGCAGCTATTGTTGATGAAATCACTGGTGAGAAATTCACATGTGCGAATACATGGAA 481
 QY 641 GlnHisSerValSerArgLeuIleGlyValaProProGlyTyrValGlyHisGluGluGly 660
 Db 482 CAACACTCTGTATCAAGATTGATTGGTGACACCGAGGATGTTGGACATGAGGAGGG 541
 QY 661 GlyGlnLeuThrGluAlaValaArgArgProGlyCysValIleLeuPheAspGluVal 680
 Db 542 GGTCAATTAACAGAGCTGTAAAGAGAGAGGCCATACAGTGTGTACTCTTTGATGAAGTT 601
 QY 681 GluLysAlaHisValAlaValPheAsnThrLeuLeuGlnValLeuAspGlyArgLeu 700
 Db 602 GAGAAACACACATCTGATTTCACACTCTTCTCAAGCTTTGATGATGGAGGTTG 661
 QY 701 ThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleIleMetThrSerAsn 720
 Db 662 ACTGACGACGACAGGAGACTGTGGATTTAGAAACACTGTGATCATGATGACCTTAAC 721
 QY 721 LeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGluValAlaArg 740

FEATURES
 source

722 CTTGGTCTGAGCATCTCTGAGTGGAGCTTTTCAGGAAATGTACCATGCAAGTCTGCTGC 781
 741 AspCysValMetArgGluValArgLysHisPheArg 752
 782 GATCAGTGTAGCAGGAGTAAGAGGCAATTTCAAG 817

JUL 8

B242455

US

INITIATION

CDNA clone

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

722 CTTGGTCTGAGCATCTCTGAGTGGAGCTTTTCAGGAAATGTACCATGCAAGTCTGCTGC 781
 741 AspCysValMetArgGluValArgLysHisPheArg 752
 782 GATCAGTGTAGCAGGAGTAAGAGGCAATTTCAAG 817

JUL 8

B242455

US

INITIATION

CDNA clone

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

JUL 8

B242455

US

INITIATION

CDNA clone

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

JUL 8

B242455

US

INITIATION

CDNA clone

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

JUL 8

B242455

US

INITIATION

CDNA clone

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

B242455

/cultivar="A17"
 /db_xref="taxon:3880"
 /clone="HOXA-28N14"
 /tissue_type="3 day old seedling roots"
 /dev_stage="24 hours after treatment in the dark at 26 C
 with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
 presence of 100 ug/ml Gentamicin"
 /lab_hosts="XLOLR"
 /clone_lib="HOXA"
 /notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
 was directionally ligated into the Unizap XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in SOLR cells."
 SE COUNT 233 a 149 c 226 g 196 t
 IGIN

Alignment Scores:
 ed. No.: 1,46e-101 Length: 804
 ore: 1169.50 Matches: 230
 cent Similarity: 93.63% Conservative: 20
 st Local Similarity: 86.14% Mismatches: 14
 ary Match: 25.70% Indels: 3
 : 14 Gaps: 1

-09-812-350-17 (1-911) x CB893763 (1-804)

481 GluGluLeuMetPheSerLeuGlnGluAlaGluArgTyrAspLeuAlaAla 500
 2 GAAGAGCTCTCTTTCACCTACAGGAGCGAGAGCGGTATGATCTAGCAAGCTGG 61
 501 AspLeuArgTyrGlyAlaLeuGlnGluValGluSerAlaLeuGlnLeuGluThr 520
 62 GACCTCGATATGTCCTCAATTAAGAGGTGGAACCTGCAATTAATAATCTTGAAGTAGC 121
 521 -----SerSerGluGlnValMetLeuThrGluAsnValGlyProGluHisIle 537
 122 ACTGATGGAAACACGGATGAGACTTATGTTGACGGAACAGTTGGACCGGACCAATA 181
 538 AlaGluValValSerArgPrpThrGlyIleProValThrArgLeuGlyGlnAsnGluLys 557
 182 GCGAGGTGTTAGCGATGAGCTGTATACAGTACTAGCTCGGCGCCAAATGAGAAA 241
 558 GluArgLeuLeuGlyLeuAlaAspArgLeuHisLysArgValValGlyGlnAsnGlnAla 577
 242 GCAAGTGTGTTGGAGCTTGGTGACAGACTGACACTAGAGTGTGCGGACCAAGACG 301
 578 ValAsnAlaValSerGluAlaIleLeuArgSerArgAlaGlyLeuGlyArgAlaGlnGln 597
 302 GTTATGCTGTGCTGAGCGCTGATTTGAGATCAGAGCTGGTTTGAAGAACCCCAACA 361
 598 ProThrGlySerPheLeuPheLeuGlyProThrGlyValGlyLysThrGluLeuAlaLys 617
 362 CCAACTGGTTCCTTCTTATCTTCTGTCGACTGGTGTGTAAACTGACCTGCGAAG 421
 618 AlaLeuAlaGluGlnPheAspAspGluAsnLeuLeuValArgIleAspMetSerGlu 637
 422 GCTCTTGAGAGCAGCTATTGATGATGAAAATCAGTTGGTGGAGAAATGACATGTCGAA 481
 638 TyrMetGluGlnHisSerValSerArgLeuIleGlyAlaProProGlyTyrValGlyHis 657
 482 TACATGGAAACAACACTCTGTATCAAGATTGATTGTCGACCAACAGGATATGTTGGACAT 541
 658 GluGluGlyGlyGlnLeuThrGluAlaValArgArgProTyrCysValIleLeuPhe 677
 542 GAGGAAGGGGTCAATTAAAGAGCTGTAAGGAGAGGCCATACAGTGTTGTACTCTTT 601
 678 AspGluValGluLysAlaHisValAlaValPheAnthrLeuLeuGlnValLeuAspAsp 697
 602 CATGAGTTGAGAAAGCACACATCTGTATTCAACACTCTTCTCCAAGTCTTGGATGAT 661

QY 698 GlyArgLeuThrAspGlyGlnGlyArgThrValAspPheArgAsnSerValIleMet 717
 Db 662 GGGAGTTGACTGACGGACAGGCGAAGCTGTGGATTTAGAAACACTGTGATCATG 721
 QY 718 ThrSerAsnLeuGlyAlaGluHisLeuLeuAlaGlyLeuThrGlyLysValThrMetGlu 737
 Db 722 ACCTCTAACCTTGGTGTGAGCATCTCTTGATGAGTCTTTCAGGAAATGATCCATGCAA 781
 QY 738 ValAlaArgAspCysValMet 744
 Db 782 GCTGCTCGCATCGAGTGATG 802
 BQ839239 740 bp mRNA linear EST 08-AUG-2002
 WHB4163_H06_O112S Wheat CS whole plant cDNA library Triticum
 aestivum cDNA clone WHB4163_H06_O11, mRNA sequence.
 BQ839239 GI:22143561
 EST.
 Triticum aestivum (bread wheat)
 Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Triticum.
 1 (bases 1 to 740)
 Anderson, O.D., Akhunov, E., Chao, S., Crossman, C., Deal, K., Dvorak, J.,
 Lazo, G.R., Rausch, C.J., Wilson, C. and Woo, J.
 The structure and function of the expressed portion of the wheat
 genomes - Chinese Spring whole plant cDNA library
 Unpublished
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Email: oanderson@pw.usda.gov
 Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: SK primer.
 Location/Qualifiers
 1. 740
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHB4163_H06_O11"
 /tissue_type="Roots, leaves, crown, stem and sheath"
 /dev_stage="Adult"
 /lab_host="E. coli SOLR"
 /clone_lib="Wheat CS whole plant cDNA library"
 /notes="Vector: Lambda Uni-ZAP XR, excised phagemid
 pBluescript SK(-); Site1: EcoRI; Site 2: XhoI; Plant
 tissues from wheat cv. CS grown to full tillering stage in
 greenhouse were collected at University of California,
 Davis (Jan Dvorak's lab, K. Deal and E. Akhunov). Total
 RNA was prepared from leaves (young leaf and third leaf),
 whole roots, crown, stem and sheath tissues, and then
 equal quantities of RNA were pooled from the these
 samples. PolyA was purified from the pooled RNA, a cDNA
 library was made, and the cDNA clones were in vivo excised
 to give pBluescript SK(-) phagemids in J. Dvorak's lab (E.
 Akhunov, J. Dvorak) at the University of California,
 Davis. Colony plating, plasmid DNA preparations and DNA
 sequencing were performed in the OD Anderson lab (all
 other authors)."
 136 a 212 c 282 g 110 t

FEATURES

source

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

 6,23e-100 Length: 740
 1152.00 Matches: 226

Percent Similarity: 98.35%
 Best Local Similarity: 93.00%
 Indels: 25.32%
 Gaps: 13

-09-812-350-17 (1-911) x BQ839239 (1-740)

153 GluSerAlaSerGlyAspThrAnPheGlnAlaLeuLysThrTrGlyArgAspLeuVal 172
 10 GAGTCCCGTCGGGGGACACCACTCCAGCGCTCAAGACGTACGGCGCGACCTCGTG 69
 173 GluGlnAlaGlyLysLeuAspProValIleGlyArgAspGluGluLeuArgValVal 192
 70 GAGGTGGCGGCAAGCTGGACCGCGTCAATCGCGCGACGAGAGATCCGGCGCGTGGTG 129
 193 ArgIleLeuSerArgThrLysAnPheGlnAlaLeuLysThrTrGlyArgAspLeuVal 212
 130 CGATCTGTGCGGGGCGCACCAAGAACCAACCCCGTCTGATCGCGGAGCCGCGTGGGC 189
 213 LysThrAlaValGluGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSer 232
 190 AGACCGCGCTGGTGGAGGGCTCGCGCGACGCGTCTGCGCGCGACGCTCCCGACCAAC 249
 233 LeuThrAspValArgLeuLysSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysTyr 252
 250 CTCCTGACGTGCGCTGCTGCGCTGGACATGGCGCGCTCTGTCGCGCGCCCAAGTAC 309
 253 ArgGlyGluPheGluGluArgLysSerValLeuLysGluValGluAlaGluGly 272
 310 CGCGCGGAGTTCGAGGAGCGGCTCAAGGCGGTCTCAAGGAGGTGGAGGAGCCGCGGG 369
 273 LysValIleLeuPheIleAspGluIleHisLeuValLeuGlyAlaGlyLysThrGluGly 292
 370 AGGTGATACTGTTATCAGCAGATACACCTGCTGGTGGCGCGCGCGGCGGAGGG 429
 293 SerMetAspAlaAlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIle 312
 430 TCCATGAGCGCGCCCAACTGTTCAAGCGGATGCTGGCGAGGGGCGACCTCAGTGCATT 489
 313 GlyAlaThrThrLeuGluGluTyrArgLysTyrValGluLysAspAlaAlaPheGluArg 332
 490 GCGCGAGCAGCCCTCGAGGAGTACAGGAAGTACGTGGAGAGGAGCGCGCTTCGAGAGG 549
 333 ArgPheGlnGlnValTyrValAlaGluProSerValProAspThrIleSerIleLeuArg 352
 550 CGGTTCAGAGGAGTGTTCGTCGAGGAGCCCGCGGTCCTGACCGCTCAGCATCTACGAGA 609
 353 GlyLeuLysGluLysTyrGluGlyHisGlyValArgIleGlnAspArgAlaLeuIle 372
 610 GGGCTCAAGGAGAGTACGAGGGGCAACCATGCTGTGAGATTTCAGGCGCGCGCTCGTC 669
 373 AsnAlaAlaGlnLeuSerAlaArgTyrIleThrGlyArgHisLeuProAspLysAlaIle 392
 670 ATTGCGCGCGAGCTCTGTCGAGGTACATCATGCTGCTCATCTGCTGATAAAGCAATC 729
 393 AspLeuVal 395
 730 GATTGGTT 738

UUT 11

Q10803

US

INITIATION

ALBEDO0002_IVaF_D06 5', mRNA

EST 16-MAY-2003

ALBEDO0002_IVaF_D06 5', mRNA sequence.

CB610803

EST

CB610803.1 GI:29550416

EST

WORDS

ORCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Sapindales; Rutaceae; Citrus.

1 (bases 1 to 722)

AUTHORS

TITLE

JOURNAL

COMMENT

Uratsu, S.L., Fisk, H.J. and Dandekar, A.M.
 Analysis of peel specific genes in Citrus
 Unpublished

Contact: Abhaya Dandekar, PhD

CAES Genome Facility
 UC Davis, Department of Pomology
 One Shields Ave, Davis, CA 95616, USA

Tel: 530 752 7784

Fax: 530 752 8502

Email: amdandekar@ucdavis.edu

Seq primer: TCCGAGATCTGACGAGC.

Location/Qualifiers

1..722

/organism="Citrus sinensis"

/mol_type="mRNA"

/cultivar="Washington navel orange"

/db_xref="taxon:2711"

/clones="ALBEDO0002_IVaF_D06"

/sex="Hermaphrodite"

/cell_type="albedo"

/dev_stage="Mature fruit sample - collected January 2002"

/lab_host="XL10-Gold"

/clone_lib="Mature Albedo"

/note="Organ: Peel/rind; Vector: pTriplEx2; Site 1: EcoRI;

Site 2: XbaI; Mature citrus fruits were harvested from

trees growing in the 'Citrus variety collection' in the

Wolfkill experimental orchard located in Winters,

California (USA). Fruit was collected in January 2002,

between 1 and 3 PM and stored at 4C. Albedo tissue was

separated from the rest of the peel and used to isolate

RNA using Trizol reagent from Invitrogen. The cDNA library

was constructed using the SMART cDNA library kit (Clontech

). The amplified library was en masse evicted and plasmid

DNA containing the cDNA library was isolated from the

resultant bacterial population. Plasmid DNA was then

transformed into ultra competent E coli cells (XL10 Gold;

Stratagene). Transformants were plated out on Q-trays

(2000 cfu/tray), picked using a Qbot and archived in 384

well dishes."

BASE COUNT 202 a 114 C 213 G 189 T 4 others

ORIGIN

Alignment Scores:

Pred. No.: 2,09e-96 Length: 722

Score: 1115.00 Matches: 220

Percent Similarity: 96.65% Conservative: 11

Best Local Similarity: 92.05% Mismatches: 8

Query Match: 24.51% Indels: 0

DB: 14 Gaps: 0

US-09-812-350-17 (1-911) x CB610803 (1-722)

QY 186 GluGluIleArgArgValValArgIleLeuSerArgArgThrLysAsnAsnProValLeu 205

Db 4 GAAGAATTAGAACGTTTGTTCGAATTCCTCAAGGAGACTATAAACAATCTCTCTT 63

QY 206 IleGlyGluProGlyValGlyLysThrAlaValValGlyLeuAlaGlnArgIleVal 225

Db 64 ATTGGAGAGCCAGGAGTTGTAAACCCCTGTGGTGGAGAGATTGGCTCAGAGAATTGTG 123

QY 226 LysGlyAspValProAsnSerLeuThrAspValArgLeuIleSerLeuAspMetGlyAla 245

Db 124 CGAGGAGATGTCTAGCAACCTTCTGATGTGAGGCTCATTTGCTTGGATATGGTCT 183

QY 246 LeuValAlaGlyAlaIleValLysValIleLeuPheIleAspGluLysSerValLeuLys 265

Db 184 CTGGTTGAGGCGCCCAAGTATAGGCGAGAGTTTGAAGAGAGGTTGAGGCTGTTCTGAAA 243

QY 266 GluValGluAspAlaGluGlyLysValIleLeuPheIleAspGluLysSerValLeuLys 285

Db 244 GAAGTAGAGAGAGCTGAGGAGAAAGTATTTTATTTATGAGAGATTCACTTGTCTT 303

QY 286 GlyAlaGlyLysThrGluGlySerMetAspAlaAlaAsnLeuPheLysProMetLeuAla 305

was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in SOLR cells."

BASE COUNT 201 a 88 c 219 g 228 t
IGIN
Alignment Scores:
ed. No.: 5.18e-96 Length: 736
ore: 1111.00 Matches: 224
Percent Similarity: 96.73% Conservative: 13
Best Local Similarity: 91.43% Mismatches: 8
ery Match: 24.42% Indels: 1
Gaps: 0

-09-812-350-17 (1-911) x BG648385 (1-736)

```

97 GlnAlaAlaGlnLysSerArgGlyAspThrHisLeuAlaValAspGlnLeuMetGly 116
|||||
1 CAAGCTGCTCAGAAATTA-CGTGGTGATCTCAATTTAGTCTGTTGATCAAGTTTGGGA 59
LeuLeuGluAspSerGlnIleArgAspLeuAsnGluValGlyValAlaThrAlaArg 136
|||||
60 ATTCTTGAGGATCTCAGATTGCTGTTGTTTAAAGAGCTGGTGTCTGTTCTAGG 119
137 ValLysSerGluValGluLysLeuArgGlyGlyGlyGlyLysValGluSerAlaSer 156
|||||
120 GTTAAAGCTGAGGTGAAAGAGCTGAGGGTAAAGATGGGAAAGGTTGAAAGTCTTC 179
157 GlyAspThrAsnGlnAlaLeuLysThrTyArgAspLeuValGluGlnAlaGly 176
|||||
180 GGTGATACGAATTTTCAAGCATTTGAAGACTTATGTTAGAGATTGTTGAAACAGCAGG 239
177 LysLeuAspProValIleGlyArgAspGluLysLeuArgValValArgIleLeuSer 196
|||||
240 AAGCTTGATCTGTTATTGACGCGCAGAGAGATTCAGAGATTTGAGATTGTCA 299
197 ArgArgThrLysAsnAsnProValLeuLysGlyGlyGlyProGlyValGlyLysThrAlaVal 216
|||||
300 AGGAGGACTAAGAAATATCCAGTCTTATTGTTGAGCTGTTGAGGAACTGCTGTT 359
217 ValGluGlyLeuAlaGlnArgIleValLysGlyAspValProAsnSerLeuThrAspVal 236
|||||
360 GTTCAAGGGTGGCTCAGAGATTTGAGAGGTGATGTTCTTACGATCTTCTGATGTT 419
237 ArgLeuLeuSerLeuAspMetGlyAlaLeuValAlaGlyAlaLysThrArgGlyGluPhe 256
|||||
420 AGGTTAATCTCTGGATATGGGAGCATTTGTTGCTGTTGAGAGTATAGGGAGATTT 479
257 GluGluArgLeuLysSerValLeuLysGluValGluAspAlaGluGlyLysValIleLeu 276
|||||
480 GAAGAGAGGTGAAGCTGTTTGAAGAAGATTGAAGAGCTGAAGGAGGTTATCTTT 539
277 PheLeuAspGluLeuHisLeuValLeuGlyAlaGlyLysThrGluGlySerMetAspAla 296
|||||
540 TTCAATGATGAGATTCACTTGTTCTTGGAGCTGGTAGAACAGAGATCAATGGATGCT 599
297 AlaAsnLeuPheLysProMetLeuAlaArgGlyGlnLeuArgCysIleGlyAlaThrThr 316
|||||
600 GCTAATCTTTTAAAGCAATGCTTCTGCTGGACAGCTTCGATGTTGTTGTTCAACA 659
317 LeuGluGluThrArgLysThrValGluLysAspAlaPheGluArgArgPheGlnGln 336
|||||
660 CTTGAGAGATATAGGAATGTTTGAAGAAAGATGCAGCATTTGAAGAGAGGTTTCAACAG 719
337 ValTyrValAlaGlu 341
|||||
720 GTTATGTTGCTGAC 734

```

BQ862449 729 bp mRNA linear EST 14-AUG-2002

SULT 15
862449
CUS

DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM

Lactuca sativa
Lactuca sativa
Lactuca sativa
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

FEATURES
source

QGC21B24.YG.ab1 QG ABCDI lettuce salinas Lactuca sativa cDNA clone
QGC21B24, mRNA sequence.

BQ862449
BQ862449.1 GI:22247914
EST.

Lactuca sativa
Lactuca sativa
Lactuca sativa
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.

1 (bases 1 to 729)
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison,
P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenome.ucdavis.edu/

Unpublished
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659

Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
Belongs to contig QG_CA_Contig4657, see http://cgpdb.ucdavis.edu/
for details.

Plate: QGC21 row: B column: 24.

Location/Qualifiers
1..729
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QGC21B24"
/lab_host="E.coli"
/clone_lib="QG ABCDI lettuce salinas"

/note="Vector: pBRCDNA5fAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_LIB-QG ABCDI lettuce salinas
TAG_TISSUE-shoots environmental stress
TAG_SEQ=TCGACGGG"

BASE COUNT 209 a 127 c 238 g 155 t
ORIGIN

Alignment Scores:

Pred. No.: 8.04e-95 Length: 729
Score: 1098.50 Matches: 215
Percent Similarity: 96.71% Conservative: 20
Best Local Similarity: 88.48% Mismatches: 7
Query Match: 24.14% Indels: 1
DB: 13 Gaps: 1

US-09-812-350-17 (1-911) x BQ862449 (1-729)

QY 466 GluLysGluArgIleAspGluIleArgArgLeuLysGlnLysArgGluGluLeuMetPhe 485
|||||
Db 3 GAAAGAGAGAGATTGATGAAATCCGAGACTGAACAGAAAGAGGAGGAGCTTTGGTG 62
QY 486 SerLeuGlnGluAlaGluArgArgTyrAspLeuAlaAlaArgAlaAspLeuArgTyrGly 505
|||||
Db 63 GCATCTACAGAGAGCAGAAAGAGATACGATTTAGCAGAGCTGCAGATTTGAATTCGA 122
QY 506 AlaLeuGlnGluValGluSerAlaIleAlaGlnLeuGluGlyThrSerSerGluGluAsn 525
|||||

123 GCAGTTCAGAAAGTGGAAACAGCAATAGCAAACTCGAAGGAACA---ACAGATGAAAT 179
526 ValMetLeuThrGluAsnValGlyProGluHisIleAlaGluValValSerArgTrpThr 545
180 GTGATGTTGACAGAAACAGTGGACCCGATCAGATAGCTGAGGTGGTGGATCGATGACC 239
546 GlyIleProValThrArgLeuGlyGlnAsnGluLysGluArgLeuIleGlyLeuAlaAsp 565
240 GGGATCCCGGTGACTCGACTTGGAAACCAATGAGAAAGAGCGGTGATCGGGTTATCGGAT 299
566 ArgLeuHisLysArgValValGlyGlnAsnGlnAlaValAsnAlaValSerGluAlaIle 585
300 CGGTGTCATCAACGGGTCGTGGGTCAAGACGCGGTGAACGCGGTGGCTGAAGCGGTG 359
586 LeuArgSerArgAlaGlyLeuGlyArgAlaGlnGlnProThrGlySerPheLeuPheLeu 605
360 CTGAGGTACAGAGCCGGCTCGGAAGGGCTCAGCAGCCACCGGATCCTTCCTGTTCTG 419
606 GlyProThrGlyValGlyLysThrGluLeuAlaLysAlaLeuAlaGluGlnLeuPheAsp 625
420 GGTCCACAGGTGTTGGAAAAACAGAGCTTGCAAAGGCTCTTCAGAGCAGTTGTTTGTAT 479
626 AspGluAsnLeuLeuValArgIleAspMetSerGluTyrMetGluGlnHisSerValSer 645
480 GATGAGAAAGCTGATGATTAGATCGATATGCGAGTATATGAGAGCAGCATTCGGTGTC 539
646 ArgLeuIleGlyAlaProGlyTyrValGlyHisGluGluGlyGlyGlnLeuThrGlu 665
540 CGGCTGATCGAGCTCCGCCAGGGTATGTTGGTCAAGAGCGCGGCGAGTTGACCGAG 599
666 AlaValArgArgProTyrCysValIleLeuPheAspGluValGluLysAlaHisVal 685
600 GCTGTGAGAGGAGACCATACAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 659
686 AlaValPheAsnThrLeuGlnValLeuAspAspGlyValArgLeuThrAspGlyGlnGly 705
660 TCGGTGTTCAATACGCTCTTCAGATGTTGATGATGAGAGGTTGACCGATGGTCAAGGT 719
706 ArgThrVal 708
720 CGAAGCGTT 728

Search completed: February 13, 2004, 04:58:38
time : 3857 secs

